

# Computation of ESR spectra from the time evolution of the magnetization: comparison of autocorrelation and Wiener-Khinchin-relation based methods

Hiroki Ikeuchi<sup>1,2,\*</sup>, Hans De Raedt<sup>3</sup>, Sylvain Bertina<sup>4</sup>, and Seiji Miyashita<sup>1,2</sup>

<sup>1</sup>*Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Bunkyo-Ku, Tokyo, 113-0033, Japan*

<sup>2</sup>*CREST, JST, 4-1-8 Honcho Kawaguchi, Saitama, 332-0012, Japan*

<sup>3</sup>*Department of Applied Physics, Zernike Institute for Advanced Materials,*

*University of Groningen, Nijenborgh 4, NL-9747AG Groningen, The Netherlands*

<sup>4</sup>*Aix-Marseille Université, CNRS, IM2NP UMR7334, F-13397 Marseille Cedex 20, France*

(Dated: August 7, 2015)

The calculation of finite temperature ESR spectra for concrete specified crystal configurations is a very important issue in the study of quantum spin systems. Although direct evaluation of the Kubo formula by means of numerical diagonalization yields exact results, memory and CPU-time restrictions limit the applicability of this approach to small system sizes. Methods based on the time evolution of a single pure quantum state can be used to study larger systems. One such method exploits the property that the expectation value of the autocorrelation function obtained for a few samples of so-called thermal typical states yields a good estimate of the thermal equilibrium value. In this paper, we propose a new method based on a Wiener-Khinchin-like theorem for quantum system. By comparison with exact diagonalization results, it is shown that both methods yield correct results. As the Wiener-Khinchin-based method involves sampling over thermal typical states, we study the statistical properties of the sampling distribution. Effects due to finite observation time are investigated and found to be different for the two methods but it is also found that for both methods, the effects vanish as the system size increases. We present ESR spectra of the one-dimensional XXZ Heisenberg chain of up to 26 spins show that double peak structure due to the anisotropy is a robust feature of these spectra.

PACS numbers: 05.30.-d, 75.10.jm, 76.30.-v

Keywords: ESR, Wiener-Khinchin

---

\* Corresponding author. Email address: ikeuchi@spin.phys.s.u-tokyo.ac.jp

## I. INTRODUCTION

Quantum spin systems have attracted interests for decades because they exhibit various nontrivial behavior due to quantum mechanical effects. In particular, in low dimensions, quantum fluctuations due to non-commutativity of the spin operators and/or competition among the interaction (frustration) play an important role and various novel concepts, such as the valence-bond solid, resonating-valence bonds, and magnon Bose-Einstein condensation, etc. have been developed. One important topic is the effect of nonmagnetic defects. In a spin  $S = 1/2$  antiferromagnetic Heisenberg chain, quantum fluctuations prevent the spins from being ordered, even at  $T = 0\text{K}$ , and the ground state is nonmagnetic. However, nonmagnetic defects break the translational symmetry and polarize the surrounding spins [1–3]. Then the one-dimensional system is described by an open-ended spin chain. An example of such a system is the Pd doped chain  $\text{Sr}_2\text{CuO}_3$  [4–6].

ESR is one of the major tools to study the effects of defects in spin systems. Modeling the ESR spectra of intrinsic defects in spin chains is an important problem for which data for finite but rather long chains are necessary. In particular, the parameter dependence of concrete ESR spectrum for a specified system is of great interest and the temperature dependence of the ESR spectrum provides a lot of information about the spin ordering. To study these aspects theoretically, the explicit form of interactions and spatial configuration of magnetic ions in the lattice play an important role and it is necessary to study microscopic models, that is we should calculate the ESR spectrum for specific quantum spin Hamiltonian. The most direct way is to calculate the Kubo formula [7, 8] by making use of the eigenvalues and their eigenvectors obtained by diagonalization of the Hamiltonian. The first attempt has been made to study the Nagata-Tazuke phenomena [9] in a one-dimensional Heisenberg chain of 8 spins with dipole-dipole interactions [10]. In this work, the dependence of the spectrum on the angle between the static field and lattice direction was reproduced. Moreover, antiferromagnetic resonance [11] and also the appearance of a resonance forbidden by the Dzyaloshinskii-Moriya interaction [12] were studied. The structure of the ESR spectrum of a spin ring at high temperatures was investigated up to 16 spins [13]. Complementary, Oshikawa and Affleck have developed a field theoretical approach by making use of the exact autocorrelation of Heisenberg chain, and have successfully analyzed low temperature properties for long (infinite) chains [14].

Obviously, the application of this method of exact diagonalization (ED) is limited to small systems for which we can obtain all the eigenvalues and their eigenvectors. For a system of  $N$   $S = 1/2$  spins, we need the memory of order  $2^{2N}$ . By making use of the symmetries of the system we may reduce the dimensions of the block diagonalized Hamiltonian. Even with these efforts, in practice, the system size is still limited to about 20 spins.

The restriction imposed by the exact diagonalization approach can be alleviated by computing the autocorrelation function (AC) from the time evolution of a pure state [15–17]. The ESR spectrum is obtained by Fourier transformation of the autocorrelation function of the transverse magnetization. In the AC method, a pure state evolves in time according to the time-evolution operator  $e^{-i\mathcal{H}t}$  the action of which can be computed efficiently by means the Chebyshev polynomial [18–21] or the Suzuki-Trotter-product-formula method [15, 22]. As the memory required to store one pure state or all eigenstates is of the order  $2^N$  and  $2^{2N}$ , respectively, the time evolution method allows us to study systems that are twice the size of those that can be studied by exact diagonalization. For compute thermal averages we need, in principle, to take the average over the initial states. However, it is known that the expectation value of a quantity  $A$  for a single random state  $|\Phi\rangle$  yields an estimate of the trace of  $A$ , that is  $\langle\Phi|A|\Phi\rangle \simeq \text{Tr } A$ , which becomes more accurate as the size of Hilbert space increases [22]. Because of this fact, there is no need to compute traces of matrices to obtain the thermal equilibrium average [16, 17, 22–24]. This approach has been used to study the temperature dependence of the total amplitude of ESR spectrum for the single molecular magnet  $\text{V}_{15}$ , consisting of 15  $S = 1/2$  spins [17]. Although this method to compute the ESR spectrum has considerable potential, properties of its applications to large systems have not yet been scrutinized in sufficient detail. An important issue, which we address in this paper, is that because the time-evolution method necessarily yields data for a finite time interval only, it is important to study how the ESR spectrum, that is the Fourier transform of this data, depends on the size of the time interval.

In the AC method, the spectrum is obtained from the autocorrelation function  $\langle M^x M^x(t) \rangle_{\text{eq}}$ . In ESR experiments, one measures the time evolution of the magnetization  $M^x(t)$ . The relation between the AC and the spectral density  $M^x(t)$  is given by the Wiener-Khinchin (WK) theorem. This theorem relates the spectral density of the dynamics of a quantity and the Fourier transform of the autocorrelation function of the same quantity. In the present paper, we propose a method to directly compute the ESR spectrum from the time evolution of  $M^x(t)$  by exploiting the idea of WK theorem. In quantum systems, however, the definition of the magnetization dynamics is somewhat tricky and we therefore develop a quantum version of Wiener-Khinchin relation, i.e., an explicit relation between Fourier transform of the autocorrelation function and the spectrum density in the quantum case. We call the approach based of this idea the Wiener-Khinchin (WK) method. Because the thermal average of the magnetization in the transverse direction  $\langle M^x(t) \rangle_{\text{eq}}$  is zero, and the ESR signal is proportional to the average of square of the Fourier transform of  $\langle\Phi|M^x(t)|\Phi\rangle$  over many realizations of the random state  $|\Phi\rangle$ , the reasoning about the convergence of this average as a function of the dimension of the Hilbert space does not hold. Therefore, we present a detailed analysis of the distribution of the sampled data, and show that the distribution converges with a finite variance, independent of the size of Hilbert space. Because of the latter property, also for a large system, it is necessary to perform ensemble averaging of the data which renders the computational efficiency of the WK method less than the one of the AC method. However, the WK method gives additional information about effects of finiteness of time domain. In particular, the Gibbs oscillations in the WK method are positive only whereas they can

be negative in the AC method. We find that as the size of system increases, for both methods, the effects of finite time interval disappear and the ESR spectra are very similar. For large systems, the AC method is the most efficient one because a single sample of a random state yields accurate result. We use both methods for applications to large systems which have not yet been studied in previous works. In particular, we compute spectra for one-dimensional XXZ model up to 26 spins and show that the double peak which was found in earlier work [13] is almost independent of the system size, indicating that the double-peak structure will be present in the thermodynamic limit.

The outline of this paper is as follows. A brief overview on the methods previously used is given in Sec II. In Sec. III, we introduce the new method motivated by the Wiener-Khinchin theorem, and study statistical properties of the method in detail in Sec. III C. In Sec. IV, we show the distribution of sampled data of the spectrum obtained by the AC and WK methods. The total amplitude of the spectrum, which is fundamental information in ESR studies, is discussed in Sec. V. Section VI presents applications of the methods to large systems. Summary and discussion of related problems are given in Sec. VII. Appendix A discusses the effect of the finite time interval on the spectrum for the both methods and Appendix B gives a detailed analysis of the statistical properties of the thermal typical state for the both methods.

## II. RECAPITULATION OF EXISTING METHODS

### A. Kubo formula and ESR spectrum

The ESR spectrum is given by the Kubo formula[7, 8]. The imaginary part of the dynamical susceptibility  $\chi''(\omega)$  reads<sup>1</sup>

$$\chi''(\omega) = \frac{1}{2}(1 - e^{-\beta\omega}) \int_{-\infty}^{\infty} \langle M^x(0)M^x(t) \rangle_{\text{eq}} e^{-i\omega t} dt, \quad (1)$$

and the ESR absorption spectrum is given by

$$I^x(\omega) = \frac{\omega\lambda_0^2}{2} \chi''(\omega), \quad (2)$$

where  $\lambda_0$  is the amplitude of the external field, and we adopt

$$M^x(t) = e^{i\mathcal{H}t} M^x e^{-i\mathcal{H}t}, \quad M^x = \sum_{i=1}^N S_i^x, \quad (3)$$

$$\langle \cdot \rangle_{\text{eq}} = \text{Tr}[\cdot e^{-\beta\mathcal{H}}] / \text{Tr}[e^{-\beta\mathcal{H}}]. \quad (4)$$

For the numerical calculation the ESR absorption spectrum, several methods have been developed. There are essentially two types of methods: (1) Exact diagonalization method [10, 13] and (2) Time-evolution of the autocorrelation function method [16].

### B. Exact diagonalization (ED) method

The most direct calculation of the Kubo formula uses the set of eigenvalues and eigenvectors  $\{E_n, |n\rangle\}_{n=1}^D$  obtained by solving the eigenvalue problem

$$\mathcal{H}|n\rangle = E_n|n\rangle \quad (5)$$

for the hamiltonian  $\mathcal{H}$ ,  $D$  denoting the dimension of the Hilbert space. In practice, we solve Eq. (5) by numerical diagonalization. The autocorrelation function  $\langle M^x(0)M^x(t) \rangle_{\text{eq}}$  is expressed as

$$\langle M^x(0)M^x(t) \rangle_{\text{eq}} = \sum_{m,n} |\langle m|M^x|n \rangle|^2 e^{i(E_m - E_n)t - \beta E_n} / Z, \quad (6)$$

where  $Z$  is the partition function

$$Z = \sum_n e^{-\beta E_n}. \quad (7)$$

---

<sup>1</sup> Throughout the paper we denote the temperature by  $\beta^{-1}$  in order to avoid confusion with the end of the time interval  $T$ .

The Fourier transform of Eq. (6) reads

$$\int_{-\infty}^{\infty} \langle M^x(0) M^x(t) \rangle_{\text{eq}} e^{-i\omega t} dt = \sum_{m,n} |\langle m | M^x | n \rangle|^2 e^{-\beta E_n} 2\pi \delta(\omega - (E_m - E_n)) / Z, \quad (8)$$

where we used the definition

$$\delta(\omega - (E_m - E_n)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(\omega - (E_m - E_n))t} dt. \quad (9)$$

The imaginary part of the dynamical susceptibility reads

$$\chi''(\omega) \equiv \sum_{m,n} D_{m,n} \delta(\omega - \omega_{m,n}), \quad (10)$$

where

$$D_{m,n} \equiv \pi (e^{-\beta E_n} - e^{-\beta E_m}) |\langle m | M^x | n \rangle|^2 / Z, \quad \omega_{m,n} \equiv E_m - E_n. \quad (11)$$

It is sufficient to consider  $\omega_{m,n} > 0$  only since we are interested in the absorption, not the emission. Note that  $\chi''(\omega) > 0$  for  $\omega > 0$ .

In the exact diagonalization approach, the spectrum consists of a finite sum of delta functions. Therefore, to draw the spectrum, we have to represent each delta function by a regular function of certain width. For example, we may replace the delta function by a Gaussian or simply use a histogram representation. The results of the exact diagonalization are exact, usually close to machine precision, but we need to know all the eigenvalues and corresponding eigenstates. Therefore, we need memory of the order of  $D^2$  for a matrix of the size  $D = 2^N$  for systems of  $N S = 1/2$  spins. If the system has symmetry, we may reduce the size. For example if the system conserves  $M^z$ , then  $D$  is reduced to  $N C_{N/2+M^z}$ . Moreover for ESR spectra, only the uniform mode is relevant and therefore only the fully symmetrized states are necessary, which also allows a reduction of  $D$ . However, the memory limitation prevents us from studying more than  $N = 20 S = 1/2$  spins. As an illustration, we use the exact diagonalization approach to study the one-dimensional spin-1/2 XXZ model in a static magnetic field  $H$  and compute the response of the magnetization along the  $x$  axis to the AC-field  $\lambda(t)$ . The hamiltonian of the system is given by

$$\mathcal{H}_{\text{tot}} = \sum_{i=1}^N (J_x S_i^x S_{i+1}^x + J_y S_i^y S_{i+1}^y + J_z S_i^z S_{i+1}^z) - g\mu_B H \sum_{i=1}^N S_i^z + \lambda_0 \cos \omega t \sum_{i=1}^N S_i^x, \quad (12)$$

where we impose the periodic boundary condition  $S_{N+1}^\alpha = S_1^\alpha$ ,  $\alpha = x, y, z$  unless otherwise mentioned. Hereafter we adopt Kelvin as the unit of energy and we set  $g\mu_B = 1$ .

In Fig. 1, we show an example of spectrum obtained by exact diagonalization method for an antiferromagnetic Heisenberg chain with  $N = 16$  with the static field  $H = 5\text{K}$  at the temperature  $\beta^{-1} = 100\text{K}$ . The notation and the parameters of the model are set to be the same as those in the model studied in the previous study[13] ( $J_x = J_y = 1\text{K}$ ,  $J_z = 0.92\text{K}$ ).

The spectrum consisting of an ensemble of the delta peaks (Eq. (10)) is plotted as a histogram with small bins  $\Delta\omega$ . In the exact diagonalization (ED) method, we are free to choose any value of  $\Delta\omega$ . Anticipating for the comparison with the two other methods, we take  $\Delta\omega = 2\pi/T$  where  $T = dt \times n_t$ ,  $dt = 0.5$  and  $n_t = 16384$  is the number of data points.

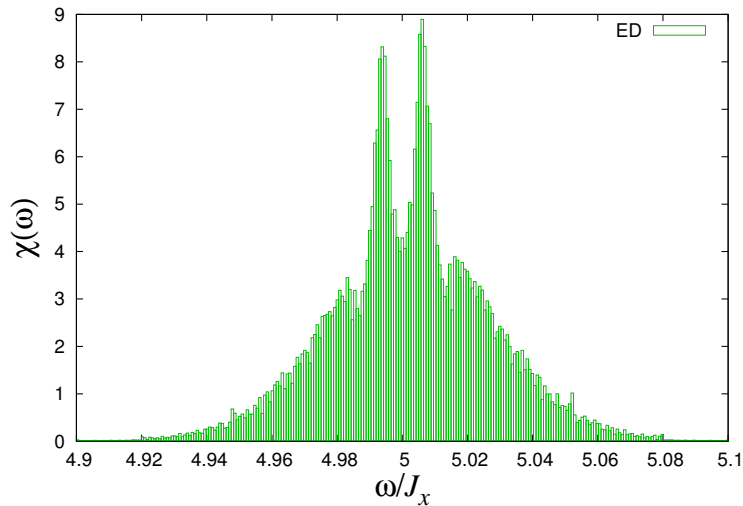


FIG. 1. Spectrum for a ring of  $N = 16$  spins. The histogram is obtained by the exact diagonalization method.

### C. Autocorrelation function method

According to the Kubo formula (Eq.(1)), the spectrum is given by Fourier transform of the autocorrelation function. Because of the symmetry

$$(\langle M^x M^x(t) \rangle_{\text{eq}})^* = \langle M^x M^x(-t) \rangle_{\text{eq}}, \quad (13)$$

where  $*$  denotes complex conjugate, we have

$$\int_{-\infty}^{\infty} \langle M^x M^x(t) \rangle_{\text{eq}} e^{-i\omega t} dt = 2\text{Re} \left[ \int_0^{\infty} \langle M^x M^x(t) \rangle_{\text{eq}} e^{-i\omega t} dt \right], \quad (14)$$

where  $\text{Re}[\cdot]$  denotes the real part. Therefore, to compute Eq. (14), it suffices to have AC data for the interval  $[0, T]$ .

We obtain the spectrum by discrete Fourier transform (DFT) of the AC data

$$f(t_m) \equiv \langle M^x M^x(t_m) \rangle_{\text{eq}}, \quad t_m = \frac{mT}{n_t}, \quad m = 0, 1, 2, \dots, n_t - 1, \quad (15)$$

where  $n_t$  is the number of data items. The DFT of Eq. (14) is given by

$$\frac{T}{n_t} \sum_{m=-n_t}^{n_t-1} f\left(\frac{mT}{n_t}\right) e^{-i\omega_k mT/n_t}, \quad \omega_k \equiv \frac{\pi k}{T}, \quad k = -n_t, -n_t + 1, \dots, 0, 1, 2, \dots, n_t - 1. \quad (16)$$

Note that the absorption and emission spectrum correspond to  $\omega_k > 0$  and  $\omega_k < 0$ , respectively. The maximum angular frequency ( $2\pi$  times Nyquist frequency) that can be represented by the DFT Eq. (16) is given by  $\pi n_t/T$ . For a magnetic field  $H = 5\text{K}$ , the main contribution to the absorption spectrum peak comes from a narrow peak around  $\omega \approx 5\text{K}$ . In our numerical work we take  $dt \equiv T/n_t = 0.5$ . Therefore the largest value of omega ( $2\pi$  times the Nyquist frequency) is  $(2\pi/T) \times (n_t/2) = \pi/dt = 2\pi \approx 6.28 > 5$ , which is large enough to cover the full spectral range.

In Fig. 2(left), we show the spectrum for a system of  $N = 6$  spins at the temperature  $\beta^{-1} = 100\text{K}$ , obtained by using a time series of  $n_t = 16384$  items. Clearly, the spectrum is suffering from fine oscillations with negative values values, which is called the Gibbs oscillation[25]. This oscillation is due to the fact that the range of the time integral is finite and that there are eigenvalues that do not not exactly match one of the  $\omega_k$ . This artifact can be suppressed by employing a suitable window function[25], for instance a Gaussian window (see Appendix A). As shown in Fig. 2 (Right), the Gibbs oscillation is reduced and the result is consistent with the one derived with the exact diagonalization method with the resolution of  $O(\alpha)$  ( $\alpha$  is an artificial parameter which determines the resolution of the spectrum. See Appendix A for the precise definition).

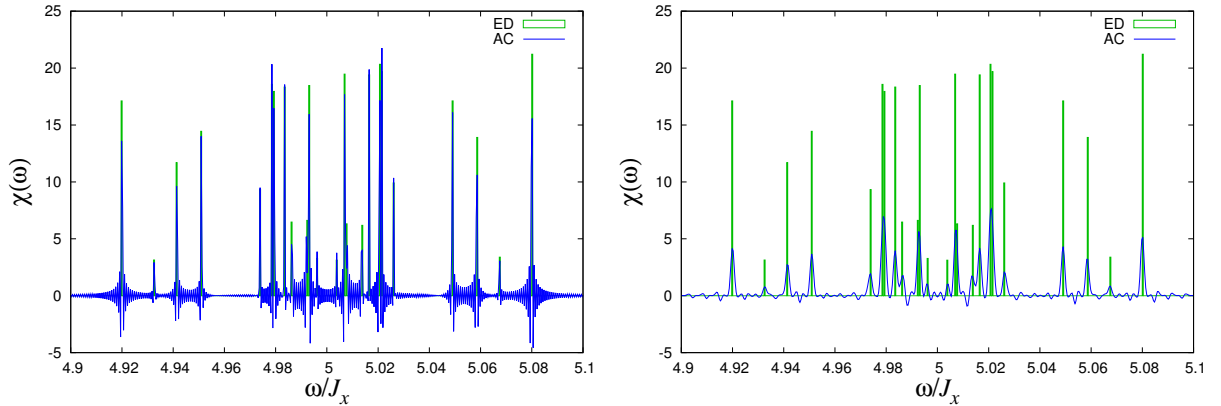


FIG. 2. The spectrum obtained by exact diagonalization method and by the time-evolution method:  $N = 6$ , temperature  $\beta^{-1} = 100\text{K}$ ,  $H = 5\text{K}$ , sampling time  $dt = 0.5$ , the number of data items  $n_t = 16384$ , and the  $\omega$ -mesh  $\pi/T = \pi/(n_t \times dt)$  (blue line). The exact diagonalization result (green line), uses an  $\omega$ -mesh  $\Delta\omega = 2\pi/16384$ , which corresponds to the frequency  $\pi/T$  used in time-evolution method. Left: the spectrum obtained by exact diagonalization (green line) and by time-evolution method without window functions (blue line). Strong Gibbs oscillation is found. Right: Same as left except that a Gaussian window function with standard deviation  $\alpha/T = 7/8192 = 0.00085$  ( $\alpha = 7$ ) is used. Clearly, the Gibbs oscillation is suppressed.

### 1. Thermal typical state method

In theory, the calculation of the thermal equilibrium expectation value of  $M^x M^x(t)$  involves a trace operation. This requires of the order  $D^2 = 2^{2N}$  operations, which becomes prohibitively large if  $N$  is of the order of 20 or larger. Fortunately, for larger systems, we can obtain accurate estimates of thermal equilibrium averages by making use of the so-called thermal typical state  $|\Phi_\beta\rangle = e^{-\beta \mathcal{H}/2} |\Phi\rangle$  [16, 17, 22, 24], also called “a Boltzmann-weighted random vector” [16, 17] or “a canonical thermal pure quantum state” [24], where  $|\Phi\rangle$  denotes a random state on the  $D$ -dimensional hypersphere. In essence, we have [22]

$$\langle X \rangle_{\text{eq}} = \frac{\text{Tr } X e^{-\beta \mathcal{H}}}{\text{Tr } e^{-\beta \mathcal{H}}} \simeq \langle \Phi_\beta | X | \Phi_\beta \rangle. \quad (17)$$

We briefly explain the ideas behind this method. Let  $\{|n\rangle\}_{n=1}^D$  be an arbitrary set of complete orthonormal states of the Hilbert space of the system. Using the complex-valued random variables  $\{\xi_n\}_{n=1}^D$ , we introduce a random vector:

$$|\Phi\rangle = \sum_{n=1}^D \xi_n |n\rangle. \quad (18)$$

Note that this construction is independent of the choice of the basis set  $\{|n\rangle\}_{n=1}^D$ .

Unlike in [A. Hams and H. De Raedt, PRE (2000)], to simplify the mathematics, we will work with independent complex-valued Gaussian random variable for each index  $n$ . They have identical Gaussian distributions with mean zero and variance  $2\sigma^2$  for all real and imaginary parts of the variables. By  $E[\cdot]$ , we denote the expectation with respect to the multivariate Gaussian probability distribution:

$$P(\xi_1, \dots, \xi_D) \prod_{a=1}^D d(\text{Re } \xi_a) d(\text{Im } \xi_a) = \prod_{a=1}^D \left[ \frac{1}{2\pi\sigma^2} e^{-|\xi_a|^2/2\sigma^2} \right] d(\text{Re } \xi_a) d(\text{Im } \xi_a). \quad (19)$$

The following relations hold:

$$\begin{aligned} E[\xi_a] &= E[\xi_p] = E[\xi_a \xi_b] = 0, \\ E[\xi_a^* \xi_p] &= 2\sigma^2 \delta_{a,p}, \\ E[\xi_a^* \xi_b^* \xi_p \xi_q] &= E[\xi_a^* \xi_p] E[\xi_b^* \xi_q] + E[\xi_a^* \xi_q] E[\xi_b^* \xi_p] = 4\sigma^4 (\delta_{a,p} \delta_{b,q} + \delta_{a,q} \delta_{b,p}), \end{aligned} \quad (20)$$

which are used in the following calculations.

For any matrix  $X$  we have

$$E[\langle \Phi | X | \Phi \rangle] = \sum_{a,p=1}^D E[\xi_a^* \xi_p] \langle a | X | p \rangle = \sum_a^D E[\xi_a^* \xi_a] \langle a | X | a \rangle = 2\sigma^2 \text{Tr } X, \quad (21)$$

and because  $\langle \Phi | X^\dagger | \Phi \rangle = \langle \Phi | X | \Phi \rangle^*$ , the corresponding variance is given by

$$\begin{aligned} \text{Var}(\langle \Phi | X | \Phi \rangle) &= E[|\langle \Phi | X | \Phi \rangle|^2] - |E[\langle \Phi | X | \Phi \rangle]|^2 \\ &= \sum_{a,p,b,q=1}^D E[\xi_a^* \xi_p \xi_b^* \xi_q] \langle a | X | p \rangle \langle b | X | q \rangle^* - 4\sigma^4 |\text{Tr } X|^2 \\ &= 4\sigma^4 \sum_{a,b=1}^D \left( \langle a | X | a \rangle \langle b | X | b \rangle^* + \langle a | X | b \rangle \langle a | X | b \rangle^* \right) - 4\sigma^4 |\text{Tr } X|^2 \\ &= 4\sigma^4 \text{Tr } X X^\dagger. \end{aligned} \quad (22)$$

The relative standard deviation is defined by

$$\text{RSD}^2(\langle \Phi | X | \Phi \rangle) = \frac{\text{Var}(\langle \Phi | X | \Phi \rangle)}{|E[\langle \Phi | X | \Phi \rangle]|^2} = \frac{\text{Tr } X X^\dagger}{|\text{Tr } X|^2} = \frac{\text{Tr } X X^\dagger}{(\text{Tr } X)(\text{Tr } X^\dagger)}. \quad (23)$$

Next, we introduce the thermal typical state  $|\Phi_\beta\rangle$  for the system at an inverse temperature  $\beta$ :

$$|\Phi_\beta\rangle = e^{-\beta \mathcal{H}/2} |\Phi\rangle, \quad (24)$$

where  $\langle \Phi_\beta | \Phi_\beta \rangle = \langle \Phi | e^{-\beta \mathcal{H}} | \Phi \rangle$  is an approximation to the partition function  $Z_\beta$ . As shown in Appendix B, the RSD of  $\langle \Phi | e^{-\beta \mathcal{H}} | \Phi \rangle$  vanishes with  $D$  as

$$\text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle) = \text{RSD}(\langle \Phi | e^{-\beta \mathcal{H}} | \Phi \rangle) \lesssim \mathcal{O}(D^{-1/2}). \quad (25)$$

Thus, hereafter we only estimate the RSD of the numerator for the thermal average. In our calculations, we take averages of the dominator and the numerator with respect to the same samples  $\{|\Phi_\beta\rangle\}$ :

$$\langle X \rangle_{\text{eq}} \simeq \frac{\frac{1}{N} \sum_{|\Phi_\beta\rangle} \langle \Phi_\beta | X | \Phi_\beta \rangle}{\frac{1}{N} \sum_{|\Phi_\beta\rangle} \langle \Phi_\beta | \Phi_\beta \rangle}, \quad (26)$$

where  $N$  is the number of samples. But if we take the average of the ratios of samples, we obtain the same result in the limit  $D \rightarrow \infty$  (see Appendix B):

$$\langle X \rangle_{\text{eq}} \simeq \frac{1}{N} \sum_{|\Phi_\beta\rangle} \frac{\langle \Phi_\beta | X | \Phi_\beta \rangle}{\langle \Phi_\beta | \Phi_\beta \rangle}. \quad (27)$$

The autocorrelation function  $\langle M^x M^x(t) \rangle_{\text{eq}}$  is given by

$$\langle M^x M^x(t) \rangle_{\text{eq}} = \frac{\text{E}[\langle \Phi_\beta | M^x e^{i\mathcal{H}t} M^x e^{-i\mathcal{H}t} | \Phi_\beta \rangle]}{\text{E}[\langle \Phi_\beta | \Phi_\beta \rangle]}. \quad (28)$$

The RSD of the autocorrelation function also vanishes with increasing  $D$  (see Appendix B). Thus, for a sufficiently large system, a few samples suffice to estimate the autocorrelation function. Here it should be noted that strictly speaking Eq. (13) does not hold for the quantity  $\langle \Phi_\beta | M^x e^{i\mathcal{H}t} M^x e^{-i\mathcal{H}t} | \Phi_\beta \rangle$ , but for sufficiently large  $D$ , Eq. (28) justifies using Eq. (13).

For the autocorrelation function  $\langle M^x M^x(t) \rangle_{\text{eq}}$ , we need to construct the states

$$|A\rangle \equiv e^{-i\mathcal{H}t} |\Phi_\beta\rangle, \quad \text{and} \quad |B\rangle \equiv e^{-i\mathcal{H}t} M^x |\Phi_\beta\rangle, \quad (29)$$

and then obtain the expectation value of the autocorrelation as

$$\langle \Phi_\beta | M^x e^{i\mathcal{H}t} M^x e^{-i\mathcal{H}t} | \Phi_\beta \rangle = \langle B | M^x | A \rangle. \quad (30)$$

To perform the operations  $e^{-\beta \mathcal{H}/2}$  and  $e^{-i\mathcal{H}t}$  on the states, we may use the Chebyshev method or the Suzuki-Trotter decomposition method. In the present study we used the Chebyshev method for  $e^{-\beta \mathcal{H}/2} |\Phi\rangle$  and mostly used the product method for  $e^{-i\mathcal{H}t} |\Phi_\beta\rangle$ .

### III. WIENER-KHINCHIN METHOD

In this section, we propose a new method that makes use of the Wiener-Khinchin theorem. In the autocorrelation method (see previous section) the spectrum is estimated from the autocorrelation function  $\langle M^x M^x(t) \rangle_{\text{eq}}$ . In experiments, the spectrum is obtained from the record of time evolution of magnetization  $M^x(t)$ . The relation between the autocorrelation and the Fourier transform of time evolution of a quantity  $X(t)$  is given by the Wiener-Khinchin relation, that is, the relation between the spectrum of the fluctuation in time of a quantity  $X(t)$  and the Fourier transform of the autocorrelation function  $\langle X(0)X(t) \rangle$ .

We explore the possibility to obtain the spectrum from the dynamics of the quantity itself, i.e.  $\langle \Phi_\beta | X(t) | \Phi_\beta \rangle$  from an initial state  $|\Phi_\beta\rangle$ . In this method, we use the time-evolution of the state as in the previous subsection and therefore we can study large systems as well.

#### A. Wiener-Khinchin theorem

First, we briefly review the Wiener-Khinchin theorem. For a time-dependent quantity  $X(t)$ , we define the autocorrelation function  $R(t)$  and the spectral density  $S(\omega)$  as

$$R(t) \equiv \langle X(0)X(t) \rangle \equiv \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(\tau)X(t+\tau) d\tau, \quad (31)$$



and

$$S(\omega) \equiv \lim_{T \rightarrow \infty} \frac{|X^T(\omega)|^2}{T}, \quad X^T(\omega) \equiv \int_0^T X(t) e^{-i\omega t} dt, \quad (32)$$

respectively. The Wiener-Khinchin theorem tells us that the Fourier transform of the autocorrelation function equals to spectral density:

$$G(\omega) \equiv \int_{-\infty}^{\infty} R(t) e^{-i\omega t} dt = S(\omega). \quad (33)$$

Note that we assume that the process  $X(t)$  is stationary[26].

### B. Dynamics of the magnetization

Next, we apply the idea of Wiener-Khinchin relation to the Kubo formula. By making use of the relation (33) we obtain  $G(\omega)$  from the spectral density  $S(\omega)$ . Here it should be noted that  $M^x(t)$  is an operator and the definition of  $\langle M^x(t) \rangle$  is tricky. Definitely,  $\langle M^x(t) \rangle_{\text{eq}}$  is time-independent, in fact zero in the present case, and therefore we cannot extract  $S(\omega)$ .

In the following, we propose a method to obtain  $G(\omega)$  from  $\langle M^x(t) \rangle$  in a quantum mechanical system. This approach is motivated by the Wiener-Khinchin theorem but we do not use the relation (33) directly. In time-domain methods, we use the notion of the thermal typical state (see above) and we use this notion here once more to derive a formula to obtain  $S(\omega)$  from time evolution of  $\langle M^x(t) \rangle$  [27].

First we prepare a thermal typical state  $|\Phi_\beta\rangle$  as an initial state. The expectation value with this state gives the thermal average at the inverse temperature  $\beta$ . For  $M^x(t)$ ,  $\langle M^x(t) \rangle_{\text{eq}} = 0$  and the expectation value  $E[\langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle]$  is zero, but, in general,  $\langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle$  is not zero. When we calculate the time evolution of the sampled state we can extract information for the spectral density

$$\hat{M}_\beta^T(\omega) \equiv \int_0^T \langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle e^{-i\omega t} dt. \quad (34)$$

Although  $E[\langle \Phi_\beta | \hat{M}^x(t) | \Phi_\beta \rangle]$  is zero, the expectation value of the spectral density  $E[|\hat{M}_\beta^T(\omega)|^2]$  is not zero and, as we show below, we can obtain the ESR spectrum from the latter quantity.

Taking the basis to be the eigenstates of the system,  $\langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle$  is expressed in terms of the random numbers  $\{\xi_n\}$  as

$$\langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle = \sum_{m,n} \xi_m^* \xi_n e^{-\beta(E_m+E_n)/2} e^{i(E_m-E_n)t} \langle m | M^x | n \rangle, \quad (35)$$

where  $|n\rangle$  and  $E_n$  are the eigenvector and its eigenenergy of the system Hamiltonian  $\mathcal{H}$  (Eq. (5)). The Fourier transform  $\hat{M}_\beta^T(\omega)$  (Eq.(34)) is expressed as

$$\hat{M}_\beta^T(\omega) = \sum_{m,n} \xi_m^* \xi_n e^{-\beta(E_m+E_n)/2} e^{-i(\omega-(E_m-E_n))T/2} 2\pi \delta^T(\omega - (E_m - E_n)) \langle m | M^x | n \rangle, \quad (36)$$

where we introduced the function

$$\delta^T(\omega) \equiv \frac{\sin \frac{\omega T}{2}}{\pi \omega} \xrightarrow{T \rightarrow \infty} \delta(\omega). \quad (37)$$

Note that we cannot extend the range of integration as in the AC method (Eq. (14)) because for a given  $|\Phi_\beta\rangle$ ,  $\langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle^* = \langle \Phi_\beta | M^x(-t) | \Phi_\beta \rangle$  does not hold in general. For the same number of time steps, this implies that the frequency resolution  $\Delta\omega = 2\pi/T$  is twice as larger as for the AC method.

As a next step, we calculate the average over the random initial states of the quantity

$$|\hat{M}_\beta^T(\omega)|^2 = \sum_{m,n} \sum_{m',n'} \xi_m^* \xi_n \xi_{m'}^* \xi_{n'} e^{-\beta(E_m+E_n)/2} e^{-\beta(E_{m'}+E_{n'})/2} \times e^{-i(\omega-(E_m-E_n))T/2} e^{+i(\omega(E_{m'}-E_{n'}))T/2} \\ \times 4\pi^2 \delta^T(\omega - (E_m - E_n)) \delta^T(\omega - (E_{m'} - E_{n'})) \langle m | M^x | n \rangle \langle n' | M^x | m' \rangle. \quad (38)$$

Using Eq. (20) and the fact that  $\langle n | M^x | n \rangle = 0$ , we have

$$E[|\hat{M}_\beta^T(\omega)|^2] = 2\sigma^2 T e^{-\beta\omega} \sum_{m,n} e^{-2\beta E_n} 2\pi \delta^T(\omega - (E_m - E_n)) |\langle m | M^x | n \rangle|^2, \quad (39)$$



where we used the relation

$$(\delta^T(\omega - (E_m - E_n)))^2 \approx \frac{T}{2\pi} \delta^T(\omega - (E_m - E_n)). \quad (40)$$

The spectral density reads

$$\Sigma_\beta(\omega) = \lim_{T \rightarrow \infty} \frac{E[|\hat{M}_\beta^T(\omega)|^2]/Z_\beta^2}{T} \quad (41)$$

$$= 2\sigma^2 e^{-\beta\omega} \sum_{m,n} e^{-2\beta E_n} 2\pi \delta(\omega - (E_m - E_n)) |\langle m|M^x|n\rangle|^2 / Z_\beta^2, \quad (42)$$

where

$$Z_\beta = E[\langle \Phi_\beta | \Phi_\beta \rangle]. \quad (43)$$

The Fourier transform of the autocorrelation function is given by

$$G_\beta(\omega) = \sum_{m,n} e^{-\beta E_n} 2\pi \delta(\omega - (E_m - E_n)) |\langle m|M^x|n\rangle|^2 / Z_\beta. \quad (44)$$

By comparing (42) and (44), we obtain a Wiener-Khinchin-like relation for the transverse magnetization:

$$G_\beta(\omega) = 2\sigma^2 \frac{Z_{\beta/2}^2}{Z_\beta} e^{\frac{\beta\omega}{2}} \Sigma_{\beta/2}(\omega). \quad (45)$$

The imaginary part of the dynamical susceptibility of interest is given by

$$\chi''(\omega) = \sigma^2 \frac{Z_{\beta/2}^2}{Z_\beta} \sinh\left(\frac{\beta\omega}{2}\right) \Sigma_{\beta/2}(\omega). \quad (46)$$

Here it should be noted that we need to calculate the quantities of  $\beta/2$  (not  $\beta$ ) to obtain the ESR spectrum of  $\beta$ .

In Fig. 3, we show a comparison of the spectrum obtained by the exact diagonalization method and by the WK method.

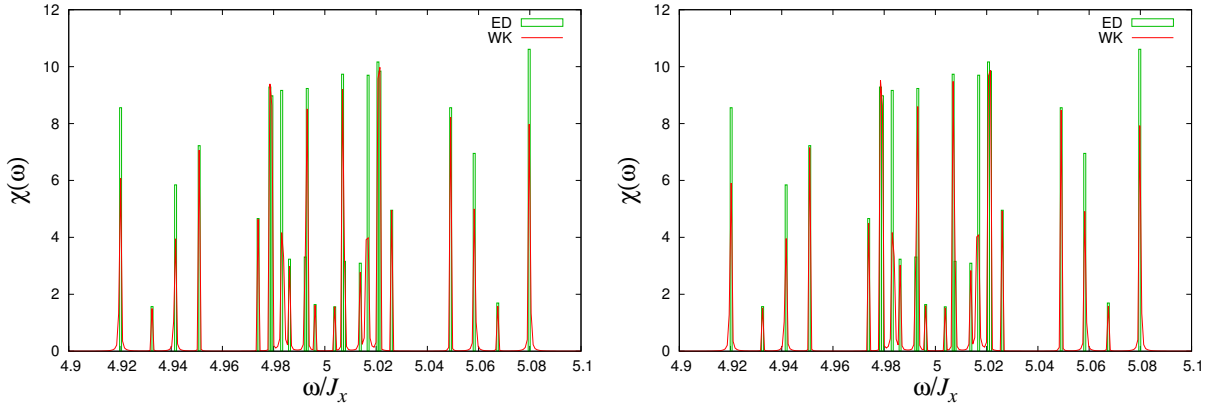


FIG. 3. Absorption spectrum obtained by the ED method (green line) and by the WK method (red line):  $N = 6$ ,  $\beta^{-1} = 100K$ ,  $H = 5K$ , and the mesh of the frequency  $\Delta\omega = 2\pi/8192$ . In the ED method (green line), we take the mesh of the frequency  $\Delta\omega = 2\pi/8192$ , which is the interval of the frequency  $2\pi/T$  in the WK method. (Left) The Wiener-Khinchin spectrum is obtained by calculating Eq. (39) exactly. (Right) The WK spectrum is obtained by averaging 10000 samples of thermal typical initial states.

From Fig. 3 it follows that the spectrum is well reproduced without any window procedure, in contrast to Fig. 2 (Left) obtained by the AC method. This fact is attributed to that the finite time effect (Gibbs oscillation) in the WK method gives positive and smaller artificial peaks in the spectrum (see the Appendix A).

### C. Sample distribution of data

In principle, methods that are based on the thermal typical state require sampling over the random states to obtain data for equilibrium. However, as mentioned above, the data obtained by the thermal typical state converge to the thermal equilibrium value as the size of system increases exponentially. That is, the deviation decreases proportionally to  $D^{-1/2}$ , where  $D$  is the dimension of the Hilbert space. In AC method, the data are obtained as expectation values of  $M^x M^x(t)$ , and for large system the convergence to the equilibrium value is so fast that one initial state suffices to obtain accurate results (see Appendix B for the details). On the other hand, the WK method extracts information from the quantity  $\hat{M}_\beta^T(\omega)$  whose expectation value is zero, and therefore we have to investigate the fluctuations of the squared variable  $|\hat{M}_\beta^T(\omega)|^2$  because it is not obvious that these fluctuations will decrease rapidly as  $D$  increases.

In this section, we study properties of sample distribution of the data. To characterize the distribution, we study the variance of  $|\hat{M}_\beta^T(\omega)|^2$ :

$$\text{Var}\left(|\hat{M}_\beta^T(\omega)|^2\right) = \text{E}\left[|\hat{M}_\beta^T(\omega)|^4\right] - \text{E}\left[|\hat{M}_\beta^T(\omega)|^2\right]^2. \quad (47)$$

To this end, we introduce

$$\tilde{M}_\beta^T(\omega) \equiv \langle \Phi | A(\omega) | \Phi \rangle, \quad (48)$$

where

$$A(\omega) = e^{-\beta \mathcal{H}/2} Y(\omega) e^{-\beta \mathcal{H}/2}, \quad (49)$$

and

$$Y(\omega) = \frac{1}{T} \int_{-T/2}^{T/2} e^{-i\omega t} e^{i\mathcal{H}t} M^x e^{-i\mathcal{H}t} dt. \quad (50)$$

which is proportional to Eq. (36), i.e.,  $\tilde{M}_\beta^T(\omega) = \hat{M}_\beta^T(\omega) \times e^{i\omega T/2}/T$ . The reason for shifting the interval from  $[0, T]$  to  $[-T/2, T/2]$  is to assure that  $Y(\omega) = Y^\dagger(-\omega)$  and

$$A_{ap} \equiv A(\omega)_{ap} = (A(-\omega)_{pa})^* = (A^\dagger(\omega)_{pa})^* = \frac{\sin[(\omega - (E_a - E_p))T/2]}{[\omega - (E_a - E_p)]T/2} \langle a | e^{-\beta \mathcal{H}/2} M^x e^{-\beta \mathcal{H}/2} | p \rangle, \quad (51)$$

where we choose the states  $|a\rangle$  to be the eigenstates of  $\mathcal{H}$ . With this choice,  $\langle a | M^x | a \rangle = 0$  and  $A_{aa} = 0$  which simplifies the calculations significantly. Using these quantities,

$$|\tilde{M}_\beta^T(\omega)|^2 = \sum_{a,b,p,q} \xi_a^* \xi_b^* \xi_p \xi_q A_{ap} A_{qb}^*. \quad (52)$$

and, using the identities Eq. (20), the expectation of Eq. (52) is expressed as

$$\text{E}[|\tilde{M}_\beta^T(\omega)|^2] = \sum_{a,b} A_{ab} A_{ab}^* = 4\sigma^4 \text{Tr} A A^\dagger. \quad (53)$$

Hereafter, for simplicity of notation but without loss of generality, we choose  $2\sigma^2 = 1$ . Note that in Eq. (53), we did not yet include other terms in (46), i.e.,  $Z_\beta$ . But, as we mentioned above, the variances of them are small and thus we only concern the variance (47).

We calculate the first term in Eq. (47) by making extensive use of the properties of Gaussian random variables. We have

$$\begin{aligned} \text{E}\left[|\tilde{M}_\beta^T(\omega)|^4\right] &= \sum_{a,b,c,d} \sum_{p,q,r,s} \text{E}\left[\xi_a^* \xi_b^* \xi_c^* \xi_d^* \xi_p \xi_q \xi_r \xi_s\right] A_{ap} A_{qb}^* A_{cr} A_{sd}^* \\ &= \sum_{a,b,c,d} \sum_{p,q,r,s} \left\{ \text{E}\left[\xi_a^* \xi_p\right] \text{E}\left[\xi_b^* \xi_c^* \xi_d^* \xi_q \xi_r \xi_s\right] + \text{E}\left[\xi_a^* \xi_q\right] \text{E}\left[\xi_b^* \xi_c^* \xi_d^* \xi_p \xi_r \xi_s\right] \right. \\ &\quad \left. + \text{E}\left[\xi_a^* \xi_r\right] \text{E}\left[\xi_b^* \xi_c^* \xi_d^* \xi_p \xi_q \xi_s\right] + \text{E}\left[\xi_a^* \xi_s\right] \text{E}\left[\xi_b^* \xi_c^* \xi_d^* \xi_p \xi_q \xi_r\right] \right\} A_{ap} A_{qb}^* A_{cr} A_{sd}^*. \end{aligned} \quad (54)$$

The first term in the curly brackets vanishes because  $\text{E}[\xi_a^* \xi_p] = \delta_{a,p}$  and  $A_{aa} = 0$ . Interchanging the summation indices  $(r, q)$  and  $(s, q)$  in the third and fourth term, respectively we obtain

$$\text{E}\left[|\tilde{M}_\beta^T(\omega)|^4\right] = 4\text{Tr} A A A^\dagger A^\dagger + |\text{Tr} A|^2 + 2(\text{Tr} A A^\dagger)^2 + 2\text{Tr} A A^\dagger A A^\dagger, \quad (55)$$

and using Eq. (53) we finally obtain for the variances

$$\text{Var} \left( |\tilde{M}_\beta^T(\omega)|^2 \right) = 4\text{Tr}AAA^\dagger A^\dagger + |\text{Tr}A^2|^2 + (\text{Tr}AA^\dagger)^2 + 2\text{Tr}AA^\dagger AA^\dagger. \quad (56)$$

We want to find bounds to the estimate of the variance Eq. (56). Let us denote the (non-negative) eigenvalues of  $AA^\dagger$  ( $A^\dagger A$ ) by  $x_k^2$  ( $y_k^2 = x_k^2$ ), that is  $AA^\dagger|x_k\rangle = x_k^2|x_k\rangle$  ( $A^\dagger A|y_k\rangle = y_k^2|y_k\rangle$ ). We assume that the eigenvectors of  $AA^\dagger$  and  $A^\dagger A$  are normalized. Then we have

$$(\text{Tr}AA^\dagger)^2 = \left( \sum_k x_k^2 \right)^2 \geq \sum_k x_k^4 = \text{Tr}AA^\dagger AA^\dagger, \quad (57)$$

$$\text{Tr}AA^\dagger A^\dagger A = \sum_{k,l} x_k^2 y_l^2 |\langle x_k | y_l \rangle|^2 \leq \sum_{k,l} x_k^2 y_l^2 = (\text{Tr}AA^\dagger) (\text{Tr}A^\dagger A) = (\text{Tr}AA^\dagger)^2. \quad (58)$$

As  $(X, Y) \equiv \text{Tr}X^\dagger Y$  defines a scalar product, by the Schwarz inequality  $|(A^\dagger, A)|^2 = |\text{Tr}A^2|^2 \leq |\text{Tr}AA^\dagger|^2$ . Noting that the last term in Eq. (56) cannot be negative, putting all this together we find

$$1 \leq 1 + \frac{4\text{Tr}AAA^\dagger A^\dagger + |\text{Tr}A^2|^2 + 2\text{Tr}AA^\dagger AA^\dagger}{(\text{Tr}AA^\dagger)^2} \leq 8. \quad (59)$$

Thus, we find the variance is bounded as

$$1 < \frac{\text{Var} \left( |\hat{M}_\beta^T(\omega)|^2 \right)}{\text{E} \left[ |\hat{M}_\beta^T(\omega)|^2 \right]^2} < 8. \quad (60)$$

indicating that in the WK method, it is necessary to average over different initial states, also for large systems. This is in sharp contrast to the case of the AC method in which the variance vanishes exponentially with the system size. But Eq. (60) also indicates that the variance is bounded, independent of the system size, and therefore we can obtain good estimates for the mean from finite samples.

We emphasize that the variance of the WK method does not depend on  $D$ , that is, the distribution shows a kind of typicality. Thus we can obtain the correct mean with aimed precision by sampling, regardless of the system size.

#### IV. DISTRIBUTION OF AVERAGE VALUE AND CONVERGENCE IN DISTRIBUTION

In this section we investigate the sample distribution of  $\chi(\omega)$  data. For concreteness, we choose the value of  $\omega$  to be 4.9916K at which the spectrum has a maximum (peak).

First we show the distribution of data in the AC method for the system  $N = 10$  where  $D = 1024$  and  $D^{-1} \approx 0.001$  is rather small. The histogram of this data is given in Fig. 4(Left). The exact value is denoted by the black line. The mean of the distribution is shown by the bold blue line. The standard deviation (square-root of the variance) is given by the arrow. Clearly Fig. 4(Left) shows that the data is distributed around the correct value with RSD=0.31025K. As we already mentioned several times, as the size of the system increases the variance will vanish as  $D^{-1}$ .

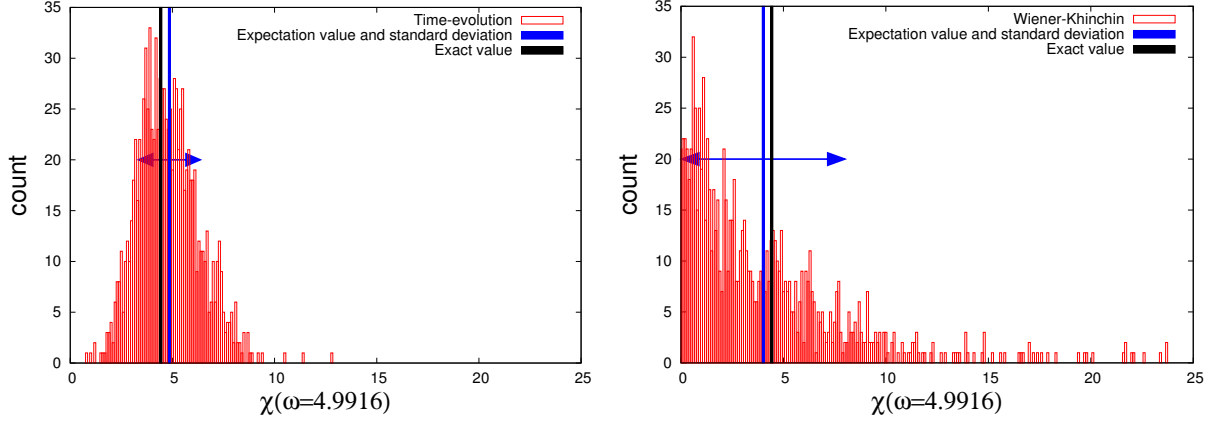


FIG. 4. (Left) The distribution of values over 1000 samples obtained by the AC method at  $\omega = 4.9916\text{K}$  for  $N = 10$ . The mean ( $=4.8508\text{K}$ ) is given by the blue line, and the standard deviation ( $=1.5049\text{K}$ ) is shown by the arrow. The exact value is denoted by the black solid line whose  $\omega = 4.4666\text{K}$ . (Right) The distribution of values over 1000 samples obtained by the WK method at  $\omega = 4.9916\text{K}$  for  $N = 10$ . The mean ( $=4.01579\text{K}$ ) is given by the blue line, and the standard deviation ( $=3.9749\text{K}$ ) is shown by the arrow. The exact value is denoted by the black solid line whose  $\omega = 4.4666\text{K}$ .

The histogram obtained by the WK method is depicted in Fig. 4(Right). In contrast to the AC method, the distribution is of the exponential type. From the figure, we find the variance is about the same as the average, i.e.,  $\text{RSD}=0.98982\text{K}$ . This value corresponds to the lower bound of the estimation (60). This fact will be discussed in more detail below.

We present the spectra for  $N = 10$  obtained by the both methods compared with the exact one obtained by the ED method. Because we have the exact eigenstates and eigenvalues, we computed the results of AC and WK methods by employing Eqs. (36) and (44) using the exact eigenstates and eigenvalues (not by sampling).

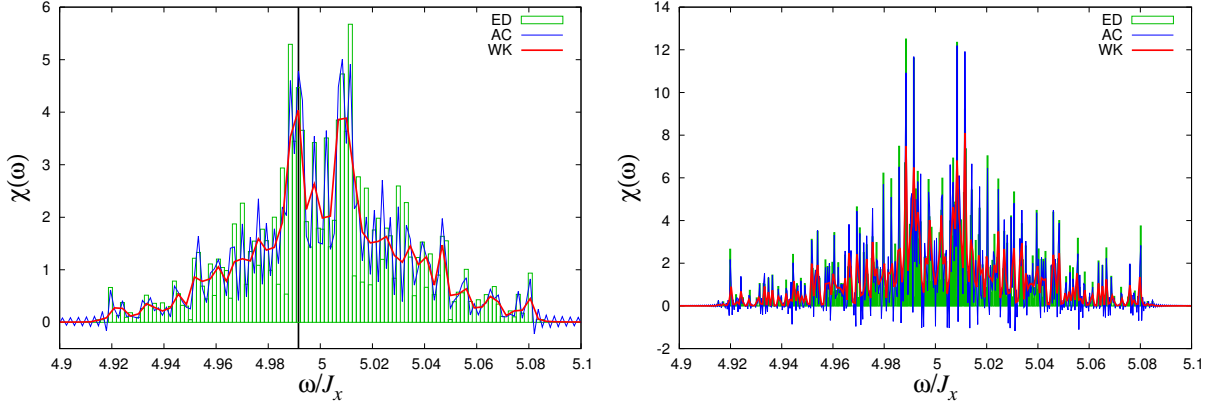


FIG. 5. (Left) Spectra obtained by the AC method (blue line) by the WK method (red line) with  $n_t = 4096$ . The column histogram is obtained by the ED method. The mesh of the frequency  $2\pi/4096$  for the WK method and  $\pi/4096$  for the AC method and the ED method. The line  $\omega = 4.9916\text{K}$  is denoted by the black line. (Right)  $n_t = 16384$ , and the mesh of the frequency  $2\pi/16384$  for the WK method and  $\pi/16384$  for the AC method and the ED method.

Although both methods approximately reproduce the exact results, the AC method shows large fluctuations for both  $n_t$  large and small. On the other hand, the WK method gives rather good agreement.

In the case of the AC method, the obtained values have a narrow distribution around the mean, and it is known that the variance decreases as  $D^{-1}$ . In this sense, the AC method has an advantage. But the WK method also gives correct estimates by sampling and the obtained shape of the spectrum is close to the exact values.

### A. Special case 1: Energy gap selection

In the previous section, we found that the relative variance is close to one, the minimum value of the bound Eq. (60). In this subsection, we consider the reason of this fact. To derive Eq. (60), we considered the general case. But we can also use the fact

that  $A_{ap}$  (Eq. (51)) is almost zero except when  $\omega \simeq E_a - E_p$ . If we assume that there is a unique set  $(n, l, m)$  which satisfies the conditions  $E_m - E_n = E_m - E_l = \omega$  and  $\langle m | M^x | n \rangle \langle m | M^x | l \rangle \neq 0$  then all terms are negligibly small except for the term  $(\text{Tr } AA^\dagger)^2$ . Therefore

$$\frac{\sqrt{\text{Var}(|\hat{M}_\beta^T(\omega)|^2)}}{\text{E}[|\hat{M}_\beta^T(\omega)|^2]} \simeq \frac{(\text{Tr } AA^\dagger)^2}{(\text{Tr } AA^\dagger)^2} = 1, \quad (61)$$

which is indeed what we have observed in our actual calculations.

### B. Special case 2: Isolated delta function

We give the variance of the WK method for the case where the spectrum consists of well-separated delta functions which are given only by a single resonant pair of states, that is the non-resonance condition that if  $E_m - E_n = E_{m'} - E_{n'}$  then  $(m = m' \text{ and } n = n')$  or  $(m = n \text{ and } m' = n')$  is satisfied. Then, we can use the fact that approximately  $\delta^T(x) = \delta(x)$  and we have

$$\text{E}[|\hat{M}_\beta^T(\omega)|^2] = 2\sigma^2 T e^{-\beta\omega} \times e^{-2\beta E_n} 2\pi \delta^T(\omega - (E_m - E_n)) |\langle m | M^x | n \rangle|^2. \quad (62)$$

The variance is estimated by making use of the relation:

$$\text{E}[\xi_m^* \xi_n \xi_{m'}^* \xi_{n'} \xi_k^* \xi_l \xi_{k'}^* \xi_{l'}] = \text{E}[(\xi_m^* \xi_m)^2 (\xi_n^* \xi_n)^2] \simeq 4 \quad (63)$$

as

$$\text{E}[|\hat{M}_\beta^T(\omega)|^4 - \text{E}[|\hat{M}_\beta^T(\omega)|^2]^2] \simeq 3\text{E}[|\hat{M}_\beta^T(\omega)|^2]^2. \quad (64)$$

Thus, the relative variance is 3, a result which does not depend on  $D$ .

In the AC method, the variance of sampled amplitude of the isolated delta functions is the same as the expectation value for the present case and  $\text{RSD} \geq 1$ . However, the isolated delta function in the spectrum is only relevant in small systems. For large system with large  $D$ , the amplitude of the isolated delta function is small and we can ignore it.

## V. THE TOTAL AMPLITUDE OF THE SPECTRUM

The total amplitude of the spectrum, or the intensity, i.e. the integral over the absorption spectrum (Eq. (2)), is given by

$$I^x = \int_0^\infty I^x(\omega) d\omega. \quad (65)$$

It is one of the fundamental pieces of information obtained in ESR experiments. Its temperature dependence has been calculated for the single molecular magnet  $\text{V}_{15}$ [17]. In this subsection, we demonstrate that  $I^x$  obtained by the WK method correctly reproduces the exact ED results. Details of the analysis are given in Appendix A. In the WK method, a delta function is represented by the sinc-function

$$\delta(\omega) \rightarrow \frac{1}{T} \left( \frac{\sin \omega T/2}{\pi \omega} \right)^2. \quad (66)$$

Because of the relation

$$\frac{1}{T} \int_{-\infty}^\infty \left( \frac{\sin \omega T/2}{\pi \omega} \right)^2 d\omega = \frac{1}{2\pi}, \quad (67)$$

the intensity  $I^x$  is obtained correctly by analytical integration. However, when the DFT yields the spectrum at discrete points  $\omega = 2k\pi/T$ ,  $k = 0, 1, 2, \dots, n_t - 1$  only. Thus, the integral over the  $\omega$  is given by a discrete sum. In Appendix A, we demonstrate that the discrete sum agrees with the analytical result. We also show that for the AC method in which the discrete points are given by  $\omega = k\pi/T$ ,  $k = -n_t, \dots, -1, 0, 1, 2, \dots, n_t - 1$  we recover the analytical result of the intensity, in spite of Gibbs oscillations. We calculate the results of AC and WK methods by evaluating the formulae for the expectation value by using the exact eigenvalues and eigenstate, as we did for Fig. 2 and 3(Left), i.e., without sampling over thermal typical states. In Fig. 6, we show that both the AC and WK methods reproduce the results of the ED method.

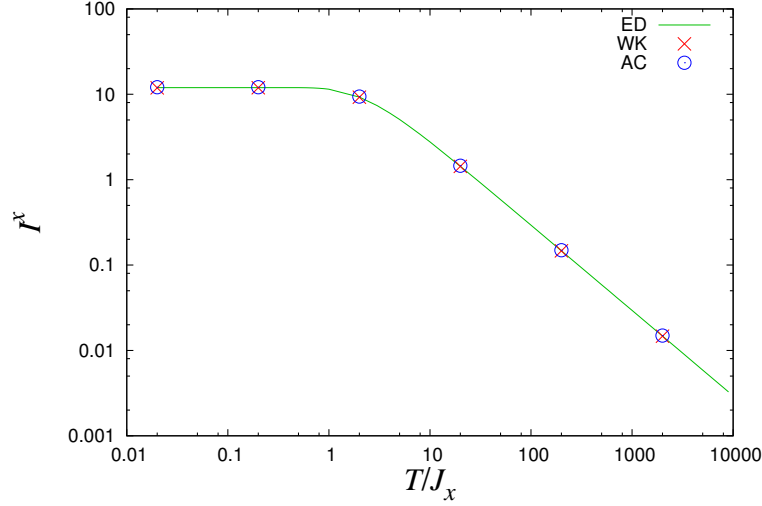


FIG. 6. Comparison of the intensities of the absorption as a function of the temperature obtained by the WK method (red cross) and by the AC method (blue open circle). The green solid curve is the exact value:  $N = 6$ ,  $H = 5K$ .

## VI. APPLICATIONS TO LARGE SYSTEMS

As an example of application of the methods to larger systems, we compute the spectrum by the AC and WK method for  $N = 16$  with  $n_t = 16384$ . For this size, we can still compute all eigenvalues and eigenstates, hence we know the positions of the eigenfrequencies and the corresponding amplitudes. The ensemble of delta functions is turned into a histogram by using a frequency mesh  $\Delta\omega = 2\pi/T$ , ( $T = 16384 \times dt (= 0.5)$ ).

In Fig. 7(Left), we present a comparison between the spectra obtained by ED and AC methods. The data obtained by the AC method shows sharp features because we did not use a window in the DFT procedure. As the statistical fluctuations on the autocorrelation are small, 10 samples suffice to find almost perfect agreement with the ED result. Therefore, we may conclude that as the size of the system increases, the effects of finiteness of the time interval reflected in e.g. Gibbs oscillations seems to be small whereas we found serious effects in the case  $N = 6, 10$ .

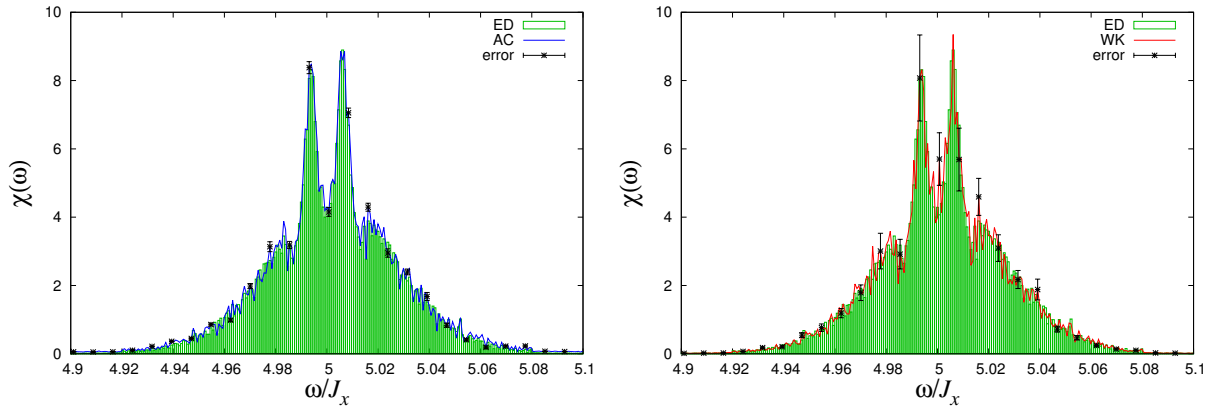


FIG. 7. (Left) Spectrum for  $N = 16$ . The histogram is obtained by exact diagonalization, points with error bars are the data obtained by the AC method with 10 samples and the time interval  $[0, T = 16384 \times dt]$ ,  $dt = 0.5$ . (Right) Results of the WK method with 52 samples over the same time interval

In Fig. 7(Right), we show the comparison between the spectra obtained by the ED and WK method. In the WK method, we took 52 samples, and we did not use a window function for the DFT procedure. Here we again find a good agreement but the variance of ensemble average is about the same as the mean as we found in the case  $N = 10$ . We conclude that for  $N = 16$ , the results of both methods are in good agreement those of ED.

With above observation, we also obtained spectra for  $N \geq 20$  by the AC method. Up to now, we adopted the periodic

boundary conditions (PBC). For a change and also because open boundary condition (OBC) is of interest because they occur in actual materials, we show results for  $N = 20$  obtained for both boundary conditions. In Fig. 8(Left), the red curve (circles) is the spectrum obtained with PBC and the blue curve with triangles is that of OBC. We find that the spectrum of OBC has a shaper double peaks in the center, but the global shape is similar in both cases. Finally, we present spectra for  $N = 24, 26$  for OBC in Fig. 8 (Center) and (Right), respectively. We used  $n_t = 16384$  with  $dt = 0.5$  for  $N = 20$ , and  $n_t = 8192$  with  $dt = 0.5$  for  $N = 24, 26$  and did not use any windowing procedure. From the size dependence, we conclude that the double peak structure will survive in the large size limit.

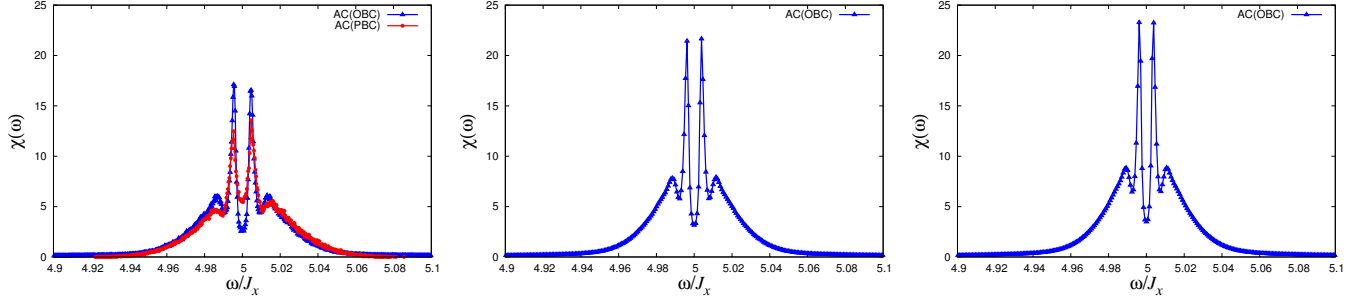


FIG. 8. Spectrum obtained by the AC method with the open boundary condition (blue line). (Left)  $N = 20$ . The red line denotes the spectrum with the periodic boundary condition. (Center)  $N = 24$ . (Right)  $N = 26$ .

## VII. SUMMARY AND DISCUSSION

We have studied time-domain methods to compute the ESR spectrum through sampling of thermal typical states. Although it has been known that the autocorrelation function can be obtained with a few samples thanks to the typicality property, practical aspects of these calculations have not yet been explored.

We proposed a method to obtain the spectrum from the motion of  $M^x$  by making use of the Wiener-Khinchin-like relation. The dynamics of magnetization from sampled initial states provides an estimate the spectrum density. But in the quantum equilibrium state the expectations value of  $M^x(t)$  and the sampled data for the spectrum density  $M_\beta^x(\omega)$  are zero. Then, we take thermal typical states as initial states, and obtain the average of squared spectrum density  $|M_\beta^x(\omega)|^2$ , i.e., the Fourier transform of the autocorrelation function. To this end, we proposed a quantum mechanical version of the Wiener-Khinchin relation. However, it should be mentioned that the WK method does not directly relate to the experimental situation. Because  $\langle M_\beta^x(\omega) \rangle_{\text{eq}} = 0$ , the averaged squared spectrum density  $|M_\beta^x(\omega)|^2$  does not converge to a mean value but has a finite distribution, also for large systems. Therefore the convergence property of the thermal typical state does not apply in this case. We studied statistical properties of the distribution and found that form of the distribution converges in the sense of typicality. We derived bounds on the relative variance of the distribution.

In time domain methods the effect of finite observation time is an important issue. We studied this aspect in detail for both the AC and WK method. It was found that for small systems, the AC method is suffering from severe effect of finiteness of  $T$  while in WK method the effect is practically suppressed. The WK and AC methods give complementary information for the effects of finite  $T$  and a comparison between them is useful to confirm that the obtained result is not affected by the finiteness observation time. We also found that for both the AC and WK method, the effect of finiteness of  $T$  decreases with increasing system size. As the system size increases, the efficiency of the AC method increases too.

We presented the spectrum for a one-dimensional XXZ chain up with to  $N = 26$  spins, where the spectrum shows a double peak which will be sustained in the thermodynamic limit. In our numerical work, we focussed on chains with an even number of spins. The ESR spectra of chains with an odd number of spins show behavior that is qualitatively different from the one of chains with an even number of spins, meriting a study in its own right. As the focus of the present paper was on time-domain methods rather than on specific applications, we relegate this study to future research.

In summary, in this paper, we proposed the new method to compute the ESR spectrum by making use of mathematical relations discussed in Sec III. We have studied one particular time-domain method based on the Wiener-Khinchin theorem but there are many other ways to obtain the spectrum density by making use of relations similar to Wiener-Khinchin theorem. For example, if we calculate  $E[\langle \Phi_\beta | M^x | \Phi_\beta \rangle \langle \Phi_\beta | M^x(t) | \Phi_\beta \rangle]$ , we can obtain a similar expression. Studying the relation among them is an interesting problem for future research.



## ACKNOWLEDGEMENTS

The present work was supported by Grants-in-Aid for Scientific Research C (25400391) from MEXT of Japan, and the Elements Strategy Initiative Center for Magnetic Materials under the outsourcing project of MEXT. The numerical calculations were supported by the supercomputer center of ISSP of Tokyo University. We also acknowledge the JSPS Core-to-Core Program: Non-equilibrium dynamics of soft matter and information.

### Appendix A: Effect of finiteness of the observation time domain and window function

#### 1. Effect of Gibbs oscillation

Because the observation time is finite  $[-T, T]$  and we have a finite number of points in time, the discrete Fourier transform (DFT) gives amplitude at the discrete points

$$\omega_k = \Delta\omega \times k, \quad \Delta\omega = \frac{\pi}{T}, \quad k : \text{integer.} \quad (\text{A1})$$

On the other hand, the resonance points are not necessarily at those points. If the spectrum has a delta-function at  $\omega$ , this is expressed by several point at the discrete points in the spectrum obtained by the DFT [25]. We study this phenomenon for the AC and WK methods.

For a time sequence of data  $\{f_k\}_{k=-\infty}^{\infty}$  such as the autocorrelation function, the spectrum of the process is given by the DFT:

$$F(\omega) = \Delta \sum_{k=-\infty}^{\infty} f_k e^{-i\omega k \Delta} \left( \simeq \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \right), \quad (\text{A2})$$

$$f_k = \frac{1}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} F(\omega) e^{i\omega k \Delta} d\omega, \quad (\text{A3})$$

where  $\Delta$  is the sampling interval. For a set with a finite number, say  $2n$ , the transformation reads

$$\tilde{F}(\omega) \equiv \Delta \sum_{k=-n}^{n-1} f_k e^{-i\omega k \Delta} = \Delta \sum_{k=-\infty}^{\infty} f_k h_k e^{-i\omega k \Delta}, \quad (\text{A4})$$

where

$$h_k = \begin{cases} 1 & k = 0, \pm 1, \pm 2, \dots, \pm(n-1), -n \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A5})$$

Expression Eq. (A4) can be written as a convolution:

$$\tilde{F}(\omega) = \frac{1}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} F(\omega') H(\omega - \omega') d\omega' = \frac{1}{2\pi} \int_{-\pi/\Delta}^{\pi/\Delta} F(\omega - \omega') H(\omega') d\omega', \quad (\text{A6})$$

where  $H(\omega)$  is the DFT of  $\{h_k\}_k$

$$H(\omega) = \Delta \sum_{k=-n}^{n-1} e^{-i\omega k \Delta} = \Delta e^{i\frac{\omega \Delta}{2}} \frac{\sin(\omega n \Delta)}{\sin(\omega \Delta / 2)}, \quad (\text{A7})$$

indicating that the spectrum we want is deformed by

$$\hat{H}(\omega) \equiv \Delta \frac{\sin(\omega n \Delta)}{\sin(\omega \Delta / 2)}. \quad (\text{A8})$$

In Fig. 9(Left), the blue dashed curve shows the modulus of Eq. (A8). If  $F(\omega)$  is a delta function  $\delta(\omega - \omega_{\text{peak}})$  the spectrum obtained by DFT is

$$\tilde{F}(\omega) = \frac{1}{2\pi} H(\omega - \omega_{\text{peak}}) \quad (\text{A9})$$

for  $\omega$  given by the mesh points  $\{\omega_k\}$ . If  $\omega_{\text{peak}}$  is one of the mesh points, the DFT spectrum has a single peak. However, if  $\omega_{\text{peak}}$  is located in an interval of mesh points, the DFT spectrum has several peaks as shown in Fig. 9(Right). The oscillation of  $H(\omega)$

even gives peaks with negative amplitude. The red rods show how the delta function in the correct spectrum at the position of green box appears in finite discrete spectrum in the AC and WK methods. In contrast to the AC method, in the WK method the spectrum is squared and the spectrum is given by  $|\tilde{F}(\omega)|^2$ , hence the spectrum is positive by construction. Although Gibbs oscillations are present, they give a width due to the finite time window and do not affect the spectrum much. In Fig. 9 and 10, we show comparison of  $\hat{H}(\omega)$  and  $|\hat{H}(\omega)|^2$  respectively. Note that in the WK method, we can obtain the discrete spectrum only at the points of even  $k$ 's.

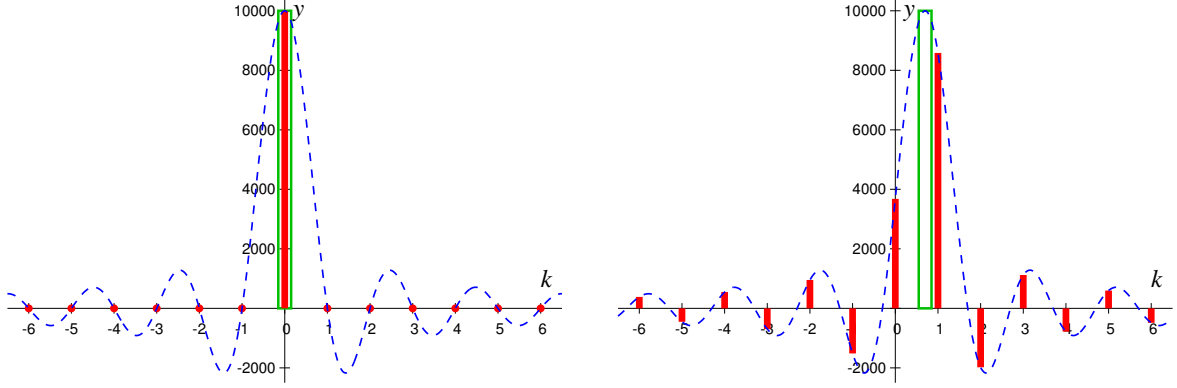


FIG. 9. the spectrum of  $|\hat{H}(\omega)|$  :  $n = 10000$ ,  $\Delta = 0.5$  : (Left) The true delta-peak is located at  $k = 0$ . (Right) The true delta-peak is located at  $k = 0.7$ .

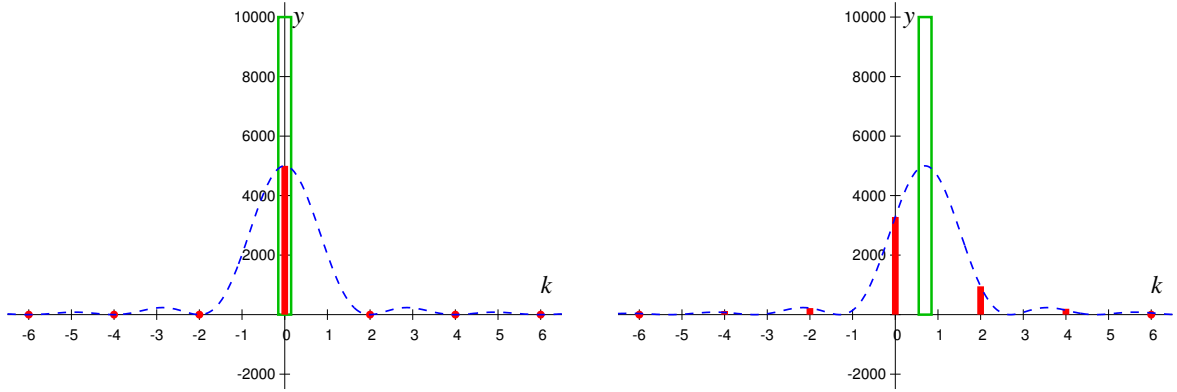


FIG. 10. the spectrum of  $\frac{1}{4T}|\hat{H}(\omega)|^2$  :  $n = 10000$ ,  $\Delta = 0.5$  : (Left) The true delta-peak is located at  $k = 0$ . (Right) The true delta-peak is located at  $k = 0.7$ .

In order to avoid these apparent negative peaks, a window function, often a Gaussian, is introduced:

$$\int_{-T}^T f(t) e^{-i\omega t} e^{-\frac{1}{2}(\alpha \frac{t}{T})^2} dt = 2\text{Re} \left[ \int_0^T f(t) e^{-i\omega t} e^{-\frac{1}{2}(\alpha \frac{t}{T})^2} dt \right]. \quad (\text{A10})$$

This treatment replaces  $f_k$  by

$$g_k = h_k e^{-\frac{1}{2}(\alpha \frac{k}{n})^2}, \quad k = 0, \pm 1, \pm 2, \dots \quad (\text{A11})$$

The resulting spectrum  $|G(\omega)|$  is depicted in Fig. 11.

The parameter  $\alpha$  determines the artificial resolution of the spectrum. Within this resolution, the Gibbs oscillations are smeared out, hence we can reproduce the spectrum as discussed in section II C.

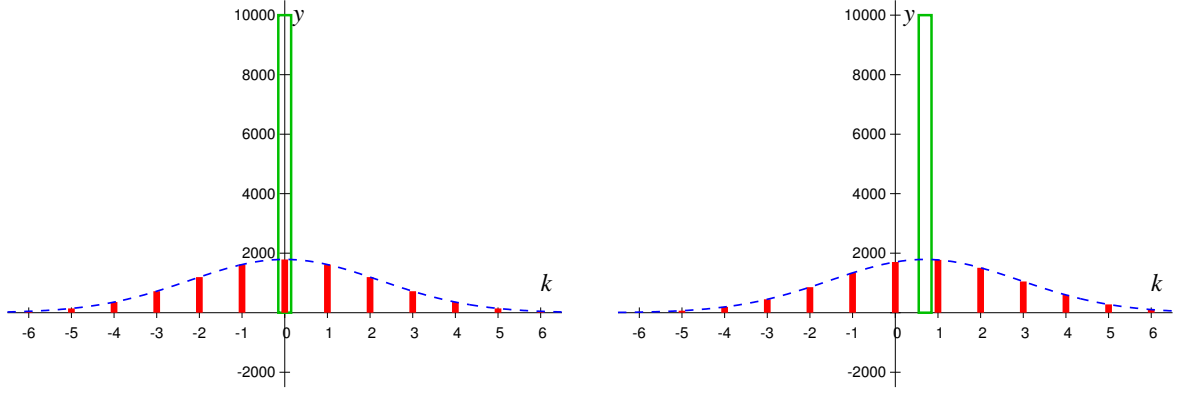


FIG. 11. The spectrum of  $|G(\omega)|$ :  $n = 10000$ ,  $\Delta = 0.5$ , and  $\alpha = 7$ . (Left) The true delta-peak is located at  $k = 0$ . (Right) The true delta-peak is located at  $k = 0.7$  (green line).

The width of the Gauss window  $\alpha$  is known to be taken as

$$e^{-\frac{1}{2}(\alpha \frac{n}{n})^2} = \varepsilon \rightarrow \alpha^2 = -2 \ln \varepsilon, \quad (\text{A12})$$

where  $\varepsilon$  is a number of the order of the smallest number of the computer resolution, say  $\varepsilon = 10^{-12}$ .

## 2. Conservation of Intensity

As mentioned in Sec. V, the total amplitude of the spectrum is reproduced correctly by the WK method and the AC method, despite the presence of Gibbs oscillations caused by performing DFT. In the subsections that follow, we demonstrate this fact analytically for both methods.

### a. Wiener-Khinchin method

In the case where  $\omega$  is a continuous variable, the total amplitude of the spectrum is independent of the time  $T$  because of the following relation;

$$\frac{1}{T} \int_{-\infty}^{\infty} \left| \frac{1}{2\pi} \int_0^T e^{-i\omega t} dt \right|^2 d\omega = \frac{1}{T} \int_{-\infty}^{\infty} \left( \frac{\sin(\frac{\omega T}{2})}{\pi \omega} \right)^2 d\omega = \frac{1}{2\pi}. \quad (\text{A13})$$

In numerical calculations, the Fourier transform is discretized and the Gibbs oscillations may seem to bring some deviation of the intensity. But in reality, as shown below, a sum rule such as Eq. (A13) holds even in the discrete case and the correct intensity can be calculated without being affected by the Gibbs oscillations, independently of the time  $T$ .

Suppose that there is a peak between two successive  $\omega_k$  values. Then the DFT and the total amplitude of the spectrum are given by

$$\frac{1}{T} \sum_{n=-M}^M \left| \frac{1}{2\pi} \sum_{k=0}^{N-1} e^{-i\omega_n k T/N} \times \frac{T}{N} \right|^2 \times \frac{2\pi}{T} = \frac{1}{2\pi N^2} \sum_{n=-M}^M \left| \frac{\sin(\frac{\omega_n T}{2})}{\sin(\frac{\omega_n T}{2N})} \right|^2, \quad (\text{A14})$$

$$\omega_n = \frac{2\pi n}{T} + \delta, \quad 0 \leq \delta < \frac{2\pi}{T}, \quad (\text{A15})$$

where  $\delta$  is the deviation of the peak's position from a discretized  $\omega$ , and  $M$  should be taken sufficiently large to sum up the amplitude around the peak. At the same time we assume that  $M \ll N$ . Then we can use the following relations:

$$\sin\left(\frac{\omega_n T}{2N}\right) = \sin\left(\frac{\pi n}{N} + \frac{T\delta}{2N}\right) \simeq \frac{\pi n}{N} + \frac{T\delta}{2N}, \quad n \in [-M, M]. \quad (\text{A16})$$

$$\sin\left(\frac{\omega_n T}{2}\right) = \sin\left(\pi n + \frac{T\delta}{2}\right) = (-1)^n \sin\left(\frac{T\delta}{2}\right). \quad (\text{A17})$$

Thus the sum is given by

$$\frac{1}{T} \sum_{n=-M}^M \left| \frac{1}{2\pi} \sum_{k=0}^{N-1} e^{-i\omega_n kT/N} \times \frac{T}{N} \right|^2 \times \frac{2\pi}{T} \simeq \frac{1}{2\pi N^2} \sum_{n=-M}^M \left( \frac{(-1)^n \sin\left(\frac{T\delta}{2}\right)}{\frac{\pi n}{N} + \frac{T\delta}{2N}} \right)^2 \simeq \frac{\sin^2\left(\frac{T\delta}{2}\right)}{2\pi^3} \sum_{n=-\infty}^{\infty} \frac{1}{\left(n + \frac{T\delta}{2\pi}\right)^2} \quad (\text{A18})$$

$$= \frac{1}{2\pi^3} \sin^2\left(\frac{T\delta}{2}\right) \times \left( \pi \operatorname{cosec}\left(\frac{T\delta}{2}\right) \right)^2 = \frac{1}{2\pi}, \quad (\text{A19})$$

where we use the formula of infinite series  $\sum_{n=-\infty}^{\infty} \frac{1}{(n+x)^2} = (\pi \operatorname{cosec}(\pi x))^2$  [28]. This result is completely in accord with the continuous one Eq. (A13).

#### b. autocorrelation method

We can also derive a similar result for the AC method. In the case where  $\omega$  is a continuous variable, the total amplitude of the spectrum is independent of the time  $T$  because of the following relation:

$$\int_{-\infty}^{\infty} 2\operatorname{Re} \left[ \int_0^T e^{-i\omega t} dt \right] d\omega = 2\pi \int_{-\infty}^{\infty} \frac{\sin(\omega T)}{\pi \omega} d\omega = 2\pi. \quad (\text{A20})$$

This sum rule holds even in the discrete case as shown below. The basic idea of the derivation is almost the same as in the WK method, but it should be noted that  $d\omega$  is half of that in the WK method because the time  $T$  is essentially two times larger in this case. The discrete version of Eq. (A20) is given by

$$\sum_{n=-M}^M 2\operatorname{Re} \left[ \sum_{k=0}^{N-1} e^{-i\omega_n kT/N} \times \frac{T}{N} \right] \times \frac{\pi}{T} = \sum_{n=-M}^M 2\operatorname{Re} \left[ \exp\left(-i\omega_n \frac{(N-1)T}{2N}\right) \right] \frac{\sin\left(\frac{\omega_n T}{2}\right)}{\sin\left(\frac{\omega_n T}{2N}\right)} \times \frac{\pi}{N} \quad (\text{A21})$$

$$= \frac{2\pi}{N} \sum_{n=-M}^M \cos\left(\omega_n \frac{(N-1)T}{2N}\right) \frac{\sin\left(\frac{\omega_n T}{2}\right)}{\sin\left(\frac{\omega_n T}{2N}\right)}, \quad (\text{A22})$$

where

$$\omega_n = \frac{\pi n}{T} + \delta, \quad 0 \leq \delta < \frac{\pi}{T}. \quad (\text{A23})$$

Using the following relations:

$$\cos\left(\omega_n \frac{(N-1)T}{2N}\right) = \cos\left(\frac{\omega_n T}{2}\right) \cos\left(\frac{\omega_n T}{2N}\right) + \sin\left(\frac{\omega_n T}{2}\right) \sin\left(\frac{\omega_n T}{2N}\right), \quad (\text{A24})$$

$$\tan\left(\frac{\omega_n T}{2N}\right) \simeq \frac{\omega_n T}{2N}, \quad n \in [-M, M], \quad \text{for } M \ll N, \quad (\text{A25})$$

and

$$\sin\left(\frac{\omega_n T}{2}\right) = \sin(\pi n + T\delta) = \sin(\pi n) \cos(T\delta) + \cos(\pi n) \sin(T\delta) = (-1)^n \sin(T\delta), \quad (\text{A26})$$

the sum is given by

$$\sum_{n=-M}^M 2\operatorname{Re} \left[ \sum_{k=0}^{N-1} e^{-i\omega_n kT/N} \times \frac{T}{N} \right] \times \frac{2\pi}{T} = \frac{2\pi}{N} \sum_{n=-M}^M \left\{ \cos\left(\frac{\omega_n T}{2}\right) \cos\left(\frac{\omega_n T}{2N}\right) + \sin\left(\frac{\omega_n T}{2}\right) \sin\left(\frac{\omega_n T}{2N}\right) \right\} \frac{\sin\left(\frac{\omega_n T}{2}\right)}{\sin\left(\frac{\omega_n T}{2N}\right)} \quad (\text{A27})$$

$$= \frac{2\pi}{N} \sum_{n=-M}^M \frac{\sin\left(\frac{\omega_n T}{2}\right) \cos\left(\frac{\omega_n T}{2}\right)}{\tan\left(\frac{\omega_n T}{2N}\right)} + \mathcal{O}\left(\frac{M}{N}\right) \simeq 2\pi \sum_{n=-M}^M \frac{\sin\left(\frac{\omega_n T}{2}\right) \cos\left(\frac{\omega_n T}{2}\right)}{\frac{\omega_n T}{2}} \quad (\text{A28})$$

$$\simeq 2\pi \sum_{n=-\infty}^{\infty} \frac{\sin\left(\frac{\omega_n T}{2}\right) \cos\left(\frac{\omega_n T}{2}\right)}{\frac{\omega_n T}{2}} = 2\pi \sum_{n=-\infty}^{\infty} \frac{\sin(\omega_n T)}{\omega_n T} \quad (\text{A29})$$

$$= 2\sin(\delta T) \sum_{n=-\infty}^{\infty} \frac{(-1)^n}{n + \frac{T\delta}{\pi}} = 2\sin(\delta T) \times \pi \operatorname{cosec}(T\delta) = 2\pi, \quad (\text{A30})$$

where we use the formula of infinite series  $\sum_{n=-\infty}^{\infty} \frac{(-1)^n}{n+x} = \pi \operatorname{cosec}(\pi x)$  [28].

## Appendix B: Statistical properties of the thermal typical state approach

The underlying idea of the thermal typical state approach is that it is a good approximation to replace  $\mathbf{Tr} X$  by  $\langle \Phi | X | \Phi \rangle$  where  $X = X^\dagger$  is a  $D \times D$  Hermitian matrix, the random state  $|\Phi\rangle = \sum_{a=1}^D \xi_a |a\rangle$  where the  $\xi_a$ 's are complex-valued Gaussian random variables and the set  $\{|a\rangle\}$  can be any complete set of orthonormal states for the Hilbert space of dimension  $D$ . The demonstration that this approximation is indeed useful requires a proof that by averaging over the  $\xi_a$  we recover the correct answer  $\mathbf{Tr} X$  and that the variance of  $\langle \Phi | X | \Phi \rangle$  is bounded.

### 1. Thermal typical state

Let us put  $X = e^{-\beta \mathcal{H}/2} Y e^{-\beta \mathcal{H}/2}$  where  $H = H^\dagger$  and  $Y = Y^\dagger$  which implies  $X = X^\dagger$ . Using Eqs. (21)–(23) we have

$$\mathbb{E}[\langle \Phi | X | \Phi \rangle] = 2\sigma^2 \mathbf{Tr} X = 2\sigma^2 \mathbf{Tr} e^{-\beta \mathcal{H}} Y, \quad (\text{B1})$$

and

$$\begin{aligned} \text{Var}(\langle \Phi | X | \Phi \rangle) &= 4\sigma^4 \mathbf{Tr} e^{-\beta \mathcal{H}/2} Y e^{-\beta \mathcal{H}/2} e^{-\beta \mathcal{H}/2} Y e^{-\beta \mathcal{H}/2} \\ &= 4\sigma^4 \mathbf{Tr} \left( e^{-\beta \mathcal{H}} Y \right) \left( e^{-\beta \mathcal{H}} Y \right) = 4\sigma^4 \mathbf{Tr} \left( e^{-\beta \mathcal{H}} Y^2 \right). \end{aligned} \quad (\text{B2})$$

In our simulations, we use the state defined by

$$|\Phi_\beta\rangle \equiv e^{-\beta \mathcal{H}/2} |\Phi\rangle, \quad (\text{B3})$$

to compute estimates of thermal equilibrium expectation values. So we can write  $\langle \Phi_\beta | Y | \Phi_\beta \rangle = \langle \Phi | X | \Phi \rangle$ . With  $A = e^{-\beta \mathcal{H}}$  and  $B = Y$  we can derive the upper bound of the variance as following;

$$\begin{aligned} \text{Var}^2(\langle \Phi_\beta | Y | \Phi_\beta \rangle) &= 16\sigma^8 |\mathbf{Tr} (AB)^2|^2 = |((AB)^\dagger, AB)|^2 \\ &\leq (AB, AB)((AB)^\dagger, (AB)^\dagger) = (\mathbf{Tr} (AB)^\dagger AB)^2 = (\mathbf{Tr} A^2 B^2)^2, \end{aligned} \quad (\text{B4})$$

where we used the Schwarz inequality because  $\mathbf{Tr} C^\dagger D \equiv (C, D)$  defines a scalar product in general.

As  $\mathbf{Tr} (AB)^2 = \mathbf{Tr} (e^{-\beta \mathcal{H}/2} Y e^{-\beta \mathcal{H}/2}) (e^{-\beta \mathcal{H}/2} Y e^{-\beta \mathcal{H}/2})^\dagger \geq 0$ , it follows that

$$\text{Var}(\langle \Phi_\beta | Y | \Phi_\beta \rangle) = 4\sigma^4 \mathbf{Tr} \left( e^{-\beta \mathcal{H}} Y \right)^2 \leq 4\sigma^4 \mathbf{Tr} e^{-2\beta \mathcal{H}} Y^2 \quad (\text{B5})$$

$$\text{RSD}^2(\langle \Phi_\beta | Y | \Phi_\beta \rangle) \leq \frac{\mathbf{Tr} e^{-2\beta \mathcal{H}} Y^2}{(\mathbf{Tr} e^{-\beta \mathcal{H}} Y)^2}. \quad (\text{B6})$$

### 2. Bounds on the variance of the estimate of the partition function

The inequality Eq. (B6) is very useful to prove that the statistical error on the estimate of the partition function vanishes exponentially with the system size. Specializing to  $Y = \mathbb{1}$  and noting that  $\langle \Phi_\beta | \Phi_\beta \rangle > 0$  we have

$$\mathbb{E}[\langle \Phi_\beta | \Phi_\beta \rangle] = 2\sigma^2 \mathbf{Tr} e^{-\beta \mathcal{H}} = 2\sigma^2 Z(\beta) = 2\sigma^2 e^{-\beta F(\beta)}, \quad (\text{B7})$$

$$\text{Var}(\langle \Phi_\beta | \Phi_\beta \rangle) = 4\sigma^4 \mathbf{Tr} e^{-2\beta \mathcal{H}} = 4\sigma^4 Z(2\beta) = 4\sigma^4 e^{-2\beta F(2\beta)}, \quad (\text{B8})$$

$$\text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle) \leq e^{-\beta(F(2\beta) - F(\beta))}, \quad (\text{B9})$$

where  $F(\beta)$  denotes the free energy of the system at the inverse temperature  $\beta$ . As the free energy is an extensive quantity, i.e. it is proportional to the number of particles, and  $F(2\beta) - F(\beta) > 0$  for  $\beta > 0$ , we have  $e^{-\beta(F(2\beta) - F(\beta))} = \mathcal{O}(e^{-N})$  and we obtain

$$\text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle) \leq e^{-\beta(F(2\beta) - F(\beta))} = \mathcal{O}(e^{-N}). \quad (\text{B10})$$

On the other hand, the justification of above estimation may not be obvious for the infinite temperature  $\beta \rightarrow 0$ . But still the convergence of the partition function is guaranteed even in this case because by using

$$\lim_{\beta \rightarrow 0} e^{-\beta F(\beta)} = \mathbf{Tr} \mathbb{1} = D, \quad (\text{B11})$$

we find that

$$\lim_{\beta \rightarrow 0} \text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle) \leq \frac{1}{\sqrt{D}}, \quad (\text{B12})$$

in agreement with [22]. Moreover, by the Schwarz inequality we have in general

$$|\text{Tr} X|^2 = |(\mathbb{1}, X)|^2 \leq (\mathbb{1}, \mathbb{1})(X, X) = D \text{Tr} X^\dagger X, \quad (\text{B13})$$

and hence

$$\text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle) = \text{RSD}(\langle \Phi | X | \Phi \rangle) \geq \frac{1}{\sqrt{D}} \quad (\text{B14})$$

from the definition of RSD given by Eq. (23). Thus, it follows that

$$\lim_{\beta \rightarrow 0} \text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle) = \frac{1}{\sqrt{D}}. \quad (\text{B15})$$

From Eqs. (B10) and (B15) we can conclude that  $\text{RSD}(\langle \Phi_\beta | \Phi_\beta \rangle)$  vanishes exponentially with the system size  $N$ . Recall that we can choose  $\sigma$  as we like. From Eq. (B7), it is clear that if we choose  $\sigma = 1/\sqrt{2}$  we have  $E[\langle \Phi_\beta | \Phi_\beta \rangle] = \text{Tr} e^{-\beta \mathcal{H}} = e^{-\beta F(\beta)}$  that is, the norm of the thermal state is, up to statistical fluctuations which vanish as  $1/D$ , equal to the partition function.

### 3. Approximate estimates

In this subsection we write  $Z = e^{-\beta \mathcal{H}}$  and to simplify the writing (but without loss of generality), we also choose  $\sigma = 1/\sqrt{2}$ . The general idea of the approach is that it suffices to generate one thermal typical state  $|\Phi_\beta\rangle$  to find good estimates for  $\langle Y \rangle_{\text{eq}} = \text{Tr} e^{-\beta \mathcal{H}} Y / \text{Tr} e^{-\beta \mathcal{H}}$ . The question is if we can prove that the statistical fluctuations are small in some sense. The problem is the following: In the simulation we generate a thermal typical state  $|\Phi_\beta\rangle$  and compute  $\langle \Phi_\beta | Y | \Phi_\beta \rangle / \langle \Phi_\beta | \Phi_\beta \rangle$ . In the present paper, we assume that the denominator and numerator are averaged and then we form the ratio. But if we take the ratio first and then we take the average, the estimate is different. In the following we study the variance in this case. Although from Eq. (B5) we know that variances of these two quantities are bounded from above, we do not yet have a bound on the variance of the ratio of them. The purpose of this subsection is to address this point.

A simple method to estimate averages and variances of the ratio is to make use of the multivariate Taylor expansion for the average

$$E\left[\frac{x}{y}\right] \approx \frac{E[x]}{E[y]} - \frac{\text{Cov}[x, y]}{E^2[y]} + \frac{E[x] \text{Var}[y]}{E^3[y]}, \quad (\text{B16})$$

where  $\text{Cov}[x, y] = E[xy] - E[x]E[y]$  and a similar approximation for the variance

$$\text{Var}\left[\frac{x}{y}\right] \approx \frac{\text{Var}[x]}{E^2[y]} - 2 \frac{E[x] \text{Cov}[x, y]}{E^3[y]} + \frac{E^2[x] \text{Var}[y]}{E^4[y]}. \quad (\text{B17})$$

Now we generate a thermal typical state  $|\Phi_\beta\rangle$  and compute ratios

$$R(Y, Z) \equiv \frac{\langle \Phi_\beta | Y | \Phi_\beta \rangle}{\langle \Phi_\beta | \Phi_\beta \rangle}. \quad (\text{B18})$$

To calculate its average and variance according to Eqs. (B16) and (B17), let us prepare the covariance  $\text{Cov}[\langle \Phi_\beta | Y | \Phi_\beta \rangle, \langle \Phi_\beta | \Phi_\beta \rangle]$ :

$$E[\langle \Phi_\beta | Y | \Phi_\beta \rangle \langle \Phi_\beta | \Phi_\beta \rangle] = \sum_{a, p, b, q=1}^D E[\xi_a^* \xi_p \xi_b^* \xi_q] \langle a | ZY | p \rangle \langle b | Z | q \rangle = \text{Tr} Z^2 Y + (\text{Tr} ZY) (\text{Tr} Z), \quad (\text{B19})$$

$$\text{Cov}[\langle \Phi_\beta | Y | \Phi_\beta \rangle, \langle \Phi_\beta | \Phi_\beta \rangle] = E[\langle \Phi_\beta | Y | \Phi_\beta \rangle \langle \Phi_\beta | \Phi_\beta \rangle] - E[\langle \Phi_\beta | Y | \Phi_\beta \rangle] E[\langle \Phi_\beta | \Phi_\beta \rangle] = \text{Tr} Z^2 Y. \quad (\text{B20})$$

By using this relation, it follows that

$$\begin{aligned} E\left[\frac{\langle \Phi_\beta | Y | \Phi_\beta \rangle}{\langle \Phi_\beta | \Phi_\beta \rangle}\right] &\approx \frac{\text{Tr} ZY}{\text{Tr} Z} - \frac{\text{Tr} Z^2 Y}{(\text{Tr} Z)^2} + \frac{\text{Tr} ZY}{\text{Tr} Z} \frac{\text{Var}[\langle \Phi_\beta | \Phi_\beta \rangle]}{(\text{Tr} Z)^2} \\ &= \frac{\text{Tr} ZY}{\text{Tr} Z} - \frac{\text{Tr} Z^2 Y}{(\text{Tr} Z)^2} + \frac{\text{Tr} ZY}{\text{Tr} Z} \frac{\text{Tr} Z^2}{(\text{Tr} Z)^2}. \end{aligned} \quad (\text{B21})$$

The second and third term in Eq. (B21) vanish exponentially with the system size because we see

$$\frac{\text{Tr } Z^2 Y}{(\text{Tr } Z)^2} \& \frac{\text{Tr } Z Y}{\text{Tr } Z} \frac{\text{Tr } Z^2}{(\text{Tr } Z)^2} \leq \|Y\| \frac{\text{Tr } Z^2}{(\text{Tr } Z)^2} = \|Y\| e^{-2\beta(F(2\beta)-F(\beta))} \quad (\text{B22})$$

where  $\|Y\| = \max_{\psi} |\langle \psi | Y | \psi \rangle|$  is the largest (in absolute value) eigenvalue of  $Y$ . Hence for large  $D$  we have

$$\mathbb{E} \left[ \frac{\langle \Phi_{\beta} | Y | \Phi_{\beta} \rangle}{\langle \Phi_{\beta} | \Phi_{\beta} \rangle} \right] \approx \frac{\text{Tr } Z Y}{\text{Tr } Z} = \langle Y \rangle_{\text{eq}}, \quad (\text{B23})$$

as expected. Similarly, for the variance we have

$$\begin{aligned} \text{Var} \left[ \frac{\langle \Phi_{\beta} | Y | \Phi_{\beta} \rangle}{\langle \Phi_{\beta} | \Phi_{\beta} \rangle} \right] &\approx \frac{\text{Var}[\langle \Phi_{\beta} | Y | \Phi_{\beta} \rangle]}{(\text{Tr } Z)^2} - 2 \frac{\text{Tr } Z Y}{\text{Tr } Z} \frac{\text{Tr } Z^2 Y}{(\text{Tr } Z)^2} + \left( \frac{\text{Tr } Z Y}{\text{Tr } Z} \right)^2 \frac{\text{Var}[\langle \Phi_{\beta} | \Phi_{\beta} \rangle]}{(\text{Tr } Z)^2} \\ &= \frac{\text{Tr } (ZY)^2}{(\text{Tr } Z)^2} - 2 \frac{\text{Tr } Z Y}{\text{Tr } Z} \frac{\text{Tr } Z^2 Y}{(\text{Tr } Z)^2} + \left( \frac{\text{Tr } Z Y}{\text{Tr } Z} \right)^2 \frac{\text{Tr } Z^2}{(\text{Tr } Z)^2} \\ &= \frac{\text{Tr } Z^2}{(\text{Tr } Z)^2} \left\{ \frac{\text{Tr } (ZY)^2}{\text{Tr } Z^2} - 2 \frac{\text{Tr } Z Y}{\text{Tr } Z} \frac{\text{Tr } Z^2 Y}{\text{Tr } Z^2} + \left( \frac{\text{Tr } Z Y}{\text{Tr } Z} \right)^2 \right\}, \end{aligned} \quad (\text{B24})$$

It is easy to find an upper bound for the terms in the curly brackets. We have

$$\left| \frac{\text{Tr } (ZY)^2}{\text{Tr } Z^2} + \left( \frac{\text{Tr } Z Y}{\text{Tr } Z} \right)^2 - 2 \frac{\text{Tr } Z Y}{\text{Tr } Z} \frac{\text{Tr } Z^2 Y}{\text{Tr } Z^2} \right| \leq 4\|Y\|^2, \quad (\text{B25})$$

and hence we find

$$\text{Var} \left[ \frac{\langle \Phi_{\beta} | Y | \Phi_{\beta} \rangle}{\langle \Phi_{\beta} | \Phi_{\beta} \rangle} \right] \leq 4\|Y\|^2 e^{-2\beta(F(2\beta)-F(\beta))}, \quad (\text{B26})$$

showing that the variance vanishes exponentially with the system size  $N$ .

- 
- [1] S. Eggert and I. Affleck, Phys. Rev. B **46**, 10866 (1992); S. Eggert and I. Affleck, Phys. Rev. Lett. **75**, 934 (1995); M. Laukamp, G. B. Martins, C. Gazza, A. L. Malvezzi, E. Dagotto, P. M. Hansen, A. C. López, and J. Riera, Phys. Rev. B **57**, 10755 (1998).
  - [2] M. Nishino, H. Onishi, K. Yamaguchi, and S. Miyashita, Phys. Rev. B **62**, 9463 (2000).
  - [3] M. Nishino, H. Onishi, P. Roos, K. Yamaguchi, and S. Miyashita, Phys. Rev. B **61**, 4033 (2000).
  - [4] M. Takigawa, N. Motoyama, H. Eisaki, and S. Uchida, Phys. Rev. B **55**, 14129 (1997).
  - [5] J. Sirker, N. Laflorencie, S. Fujimoto, S. Eggert, and I. Affleck, Phys. Rev. Lett. **98**, 137205 (2007).
  - [6] K. M. Kojima, J. Yamanobe, H. Eisaki, S. Uchida, Y. Fudamoto, I. M. Gat, M. I. Larkin, A. Savici, Y. J. Uemura, P. P. Kyriakou, M. T. Rovers, and G. M. Luke, Phys. Rev. B **70**, 094402 (2004).
  - [7] R. Kubo and K. Tomita, J. Phys. Soc. Jpn. **9**, 888 (1954).
  - [8] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957).
  - [9] K. Nagata and Y. Tazuke, **32**, 337 (1972).
  - [10] S. Miyashita, T. Yoshino and A. Ogasahara, J. Phys. Soc. Jpn. **68**, 655-661 (1999).
  - [11] A. Ogasahara and S. Miyashita, J. Phys. Soc. Jpn. **69**, 4043 (2000).
  - [12] S. Miyashita and A. Ogasahara, J. Phys. Soc. Jpn. **72**, 2350 (2003).
  - [13] S. El Shawish, O. Cépas and S. Miyashita, Phys. Rev. B **81**, 224421 (2010).
  - [14] M. Oshikawa and I. Affleck, Phys. Rev. Lett. **82**, 5136 (1999); M. Oshikawa and I. Affleck, Phys. Rev. B **65**, 134410 (2002).
  - [15] P. de Vries and H. De Raedt, Phys. Rev. B **47**, 7929 (1993).
  - [16] T. Iitaka and T. Ebisuzaki, Phys. Rev. Lett. **90**, 047203 (2003).
  - [17] M. Machida, T. Iitaka, and S. Miyashita, Phys. Rev. B **86**, 224412 (2012).
  - [18] H. Tal-Ezer and R. Kosloff, J. Chem. Phys. **81**, 3967 (1984).
  - [19] C. Leforestier, R. Bisseling, C. Cerjan, M. D. Feit, R. Friesner, A. Gulberg, A. Hammerich, G. Jolicard, W. Karrlein, H.-D. Meyer, N. Lipkin, O. Roncero, and R. Kosloff, J. Comp. Phys. **94**, 59 (1991).
  - [20] V. V. Dobrovitski and H. A. De Raedt, Phys. Rev. E **67**, 056702 (2003).
  - [21] T. Iitaka, S. Nomura, H. Hirayama, X. Zhao, Y. Aoyagi, and T. Sugano, Phys. Rev. E **56**, 1222 (1997).



- [22] A. Hams and H. De Raedt, Phys. Rev. E **62**, 4365 (2000).
- [23] H. De Raedt, A. H. Hams, K. Michielsen, S. Miyashita, and K. Saito, Prog. Theor. Phys. **138**, 489 (2000).
- [24] S. Sugiura and A. Shimizu, Phys. Rev. Lett. **108** 240401 (2012); S. Sugiura and A. Shimizu, Phys. Rev. Lett. **111** 010401 (2013).
- [25] F. J. Harris, Proc. IEEE, **66**, 51 (1978).
- [26] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed., Elsevier, Amsterdam, 2007.
- [27] The relation to real measurements is a very interesting problem which will be discussed separately.
- [28] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Corrected and enlarged edition (A. Jeffrey, ed.), Academic Press, New York, 1980.